ISSUES IN THE PHILOSOPHY OF COSMOLOGY

George F R Ellis

1 INTRODUCTION

Cosmology is the study of the large-scale structure of the Universe, where '*the Universe'* means all that exists in a physical sense [Harrison, 2000]. This is to be distinguished from *the Observable Universe*, namely that part of the Universe containing matter accessible to our astronomical observations, which is a subset of the Universe proper. Thus cosmology considers the vast domain of galaxies, clusters of galaxies, quasi-stellar objects, etc., observable in the sky by use of telescopes of all kinds, examining their nature, distribution, origins, and relation to their larger environment. *Observational cosmology* [Hoyle, 1960; Kristian and Sachs, 1966; Gunn *et al.*, 1978; Sandage *et al.*, 1993; Bothun, 1998] aims to determine the large-scale geometry of the observable universe and the distribution of matter in it from observations of radiation emitted by distant objects, while *physical cosmology* [Peebles, 1971; Sciama, 1971; Weinberg, 1972; Silk, 2001; Perkins, 2005; Dodelson, 2003] is the study of interactions during the expansion of the universe in its early hot big bang phase, and *astrophysical cosmology* [Sciama, 1971; Peebles, 1993b; Padmanbhan, 1993; Rees, 1995; Dodelson, 2003] studies the resulting later development of large-scale structures such as galaxies and clusters of galaxies. Various forms of quantum cosmology (see e.g. [Hawking, 1993; Gibbons *et al.*, 2003; Copeland *et al.*, 2005]) and studies of particle physics aspects of cosmology [Kolb and Turner, 1990; Peacock, 1999; Allday, 2002; Perkins, 2005; Dodelson, 2003] attempt to characterize the epochs before the hot big bang phase. These studies function in a mainly symbiotic way, each informing and supplementing the others to create an overall cosmological theory of the origin and evolution of the physical universe [Bondi, 1960; Harrison, 2000; Silk, 1997].

A unique role of the universe is in creating the environment in which galaxies, stars, and planets develop, thus providing a setting in which local physics and chemistry can function in a way that enables the evolution of life on planets such as the Earth. If the cosmological environment were substantially different, local conditions would be different and in most cases we would not be here [Carr and Rees, 1979; Davies, 1982; Barrow and Tipler, 1984; Tegmark, 1998; Rees, 1999] — indeed no biological evolution at all would have taken place. Thus cosmology is of substantial interest to the whole of the scientific endeavor, for it sets the

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods

framework for the rest of science, and indeed for the very existence of observers and scientists. It is unique as the ultimate historical/geographical science.

Cosmology as a serious scientific study began with the discovery of Einstein's static universe in 1917, followed by the key observational discovery of the linear redshift-distance relation by Hubble in 1929, indicating the expansion of the universe, and the development of theoretical understanding of the geometry and dynamics of the non-static Friedmann-Lemaître models with their Robertson-Walker geometry [North, 1965; Berendzen *et al.*, 1976; Smith, 1982; Ellis, 1989; Kragh, 1996]. It has been transformed in the past decades into a mainstream branch of physics [Barnett *et al.*, 1996; Nilsson *et al.*, 1991] by the linking of nuclear and particle physics theories to observable features of the cosmos [Weinberg, 1972; Kolb and Turner, 1990; Peacock, 1999; Allday, 2002; Dodelson, 2003], and into an important part of astronomy because of the massive flow of new astronomical data becoming available [Gunn *et al.*, 1978; Harwit, 1984; Bothun, 1998], particularly through new ground-based telescopes such as Keck and through balloon and satellite observatories such as the Hubble Space telescope (optical and ultraviolet), IRAS (infra-red), ROSAT (x-ray), and COBE and WMAP (microwave). Thus the subject has progressed from a mainly mathematical and even philosophical exercise to an important part of mainstream science, with a well-established standard model confirmed by various strands of evidence [Weinberg, 1972; Peebles *et al.*, 1991; Silk, 1997; Peacock, 1999; Dodelson, 2003]. Nevertheless because of its nature, it is different from any other branch of the natural sciences, its unique features playing themselves out in the ongoing interaction between speculation, theory, and observation.

Cosmology's major difference from the other sciences is the uniqueness of its object of study — the Universe as a whole [McCrea, 1953; McCrea, 1960; Munitz, 1962] — together with its role as the background for all the rest of physics and science, the resulting problems being accentuated by the vast scale of the universe and by the extreme energies occurring in the very early universe. We are unable to manipulate in any way its originating conditions, and there are limitations on our ability to observe both to very distant regions and to very early times. Additionally, there are limits to our ability to test the physics relevant at the earliest epochs. Consequently it is inevitable that (as is also the case for the other historical sciences) philosophical choices will to some degree shape the nature of cosmological theory, particularly when it moves beyond the purely descriptive to an explanatory role [Matravers *et al.*, 1995] — that move being central to its impressive progress in recent decades. These philosophical choices will strongly influence the resulting understanding, and even more so if we pursue a theory with more ambitious explanatory aims.

After a substantial outline of present day cosmology in Section 2, these issues will be explored in the subsequent sections, based on a series of thirty-four *Theses* clustered around nine key aspects of the nature of cosmology, broadly speaking relating to geometry, physics, and philosophy, that frame the context of the philosophical issues facing cosmology and its relation to local physics. I believe this formulation helps focus on specific issues of importance in this relation. To those who believe cosmology is simply about determining a number of physical parameters, this will seem a vastly over-complicated approach; but then a major purpose of this paper is precisely to counter such simplistic visions of the nature of cosmology. For other reports on the philosophy of cosmology, see [McCrea, 1970; Munitz, 1962; Ellis, 1991; Leslie, 1994; Leslie, 1998].

2 OUTLINE OF COSMOLOGY

A series of basic features of present day cosmology are now well established. Decades of painstaking work has established the distances of galaxies and hence the huge scale of the universe, as well as the basic feature that the universe is expanding and evolving; the old dream of a static universe is unviable [Ellis, 1990]. Cosmology proceeds by assuming *the laws of physics are the same everywhere, and underlie the evolution of the universe*. The dominant role of gravity, despite its weakness, then arises from the fact that it is the only known force acting effectively on astronomical scales (the other known long-range force is electromagnetism, but in this case negative charges balance out positive charges, leaving no resultant large-scale effect). Consequently, cosmological theory describing all but the very earliest times is based on the classical relativistic theory of gravitation, namely Einstein's General Theory of Relativity [Malament, 2006], with the matter present determining space-time curvature and hence the evolution of the universe. The way this works out in any particular situation depends on the nature of the matter/fields present, described by their effective equations of state and interaction potentials.

The survey of cosmology in this section looks successively at the basic models of cosmology; the hot big bang; cosmological observations, including the Cosmic Background Radiation anisotropy spectrum; causal and visual horizons, and their implications; recent theoretical developments (including inflation); the very early universe; and the present concordance model, which includes both dark matter and dark energy.

2.1 Basic Theory

Cosmology starts by assuming that *the large-scale evolution of spacetime can be determined by applying Einstein's field equations of Gravitation* ('EFE') *everywhere*: global evolution will follow from local physics. The standard models of cosmology [Robertson, 1933; Ehlers, 1993; Weinberg, 1972; Hawking and Ellis, 1973] are based on the assumption that once one has averaged over a large enough physical scale, *isotropy is observed by all fundamental observers* (the preferred family of observers associated with the average motion of matter in the universe). When this isotropy is exact, *the universe is spatially homogeneous as well as isotropic* [Walker, 1944; Ehlers, 1993; Ellis, 1971a]. The matter motion is then along irrotational and shear-free geodesic curves with tangent vector u^a, implying the existence of a canonical time-variable t obeying $u_a = -t_{,a}$. The Robertson-Walker ('RW') geometries used to describe the large-scale structure of the universe [Robertson, 1935; Walker, 1936] embody these symmetries exactly. Consequently they are conformally flat, that is, the Weyl tensor is zero:

$$(1) \quad C_{ijkl} := R_{ijkl} + \tfrac{1}{2}(R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il}) - \tfrac{1}{6}R(g_{ik}g_{jl} - g_{il}g_{jk}) = 0;$$

this tensor represents the free gravitational field, enabling non-local effects such as tidal forces and gravitational waves which do not occur in the exact RW geometries.

Comoving coordinates can be chosen so that the metric takes the form:

$$(2) \quad ds^2 = -dt^2 + S^2(t)\,d\sigma^2, \quad u^a = \delta^a{}_0 \; (a = 0, 1, 2, 3)$$

where $S(t)$ is the time-dependent scale factor, and the worldlines with tangent vector $u^a = dx^a/dt$ represent the histories of fundamental observers. The space sections $\{t = const\}$ are surfaces of homogeneity and have maximal symmetry: they are 3-spaces of constant curvature $K = k/S^2(t)$ where k is the sign of K. The normalized metric $d\sigma^2$ characterizes a 3-space of normalized constant curvature k; coordinates (r, θ, ϕ) can be chosen such that

$$(3) \quad d\sigma^2 = dr^2 + f^2(r)\left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

where $f(r) = \{\sin r,\, r,\, \sinh r\}$ if $k = \{+1,\, 0,\, -1\}$ respectively. The rate of expansion at any time t is characterised by the *Hubble parameter* $H(t) = \dot{S}/S$.

To determine the metric's evolution in time, one applies the Einstein Field Equations ('EFE'), showing the effect of matter on space-time curvature, to the metric (2,3). Because of local isotropy, the matter tensor T_{ab} necessarily takes a perfect fluid form relative to the preferred worldlines with tangent vector u^a:

$$(4) \quad T_{ab} = (\mu + p/c^2)u_a u_b + (p/c^2)g_{ab}$$

(c is the speed of light). The energy density $\mu(t)$ and pressure term $p(t)/c^2$ are the timelike and spacelike eigenvalues of T_{ab}. The integrability conditions for the EFE are the *energy-density conservation equation*

$$(5) \quad T^{ab}{}_{;b} = 0 \Leftrightarrow \dot{\mu} + (\mu + p/c^2)3\dot{S}/S = 0\,.$$

This becomes determinate when a suitable equation of state function $w := pc^2/\mu$ relates the pressure p to the energy density μ and temperature T: $p = w(\mu, T)\mu/c^2$ (w may or may not be constant). Baryons have $\{p_{bar} = 0 \Leftrightarrow w = 0\}$ and radiation has $\{p_{rad}c^2 = \mu_{rad}/3 \Leftrightarrow w = 1/3,\, \mu_{rad} = aT^4_{rad}\}$, which by (5) imply

$$(6) \quad \mu_{bar} \propto S^{-3},\; \mu_{rad} \propto S^{-4},\; T_{rad} \propto S^{-1}.$$

The scale factor $S(t)$ obeys the *Raychaudhuri equation*

$$(7) \quad 3\ddot{S}/S = -\frac{1}{2}\kappa(\mu + 3p/c^2) + \Lambda,$$

where κ is the gravitational constant and Λ the cosmological constant.[1] This shows that the active gravitational mass density of the matter and fields present is $\mu_{grav} := \mu + 3p/c^2$. For ordinary matter this will be positive:

(8) $\mu + 3p/c^2 > 0 \;\Leftrightarrow\; w > -1/3$

(the 'Strong Energy Condition'), so ordinary matter will tend to cause the universe to decelerate ($\ddot{S} < 0$). It is also apparent that a positive cosmological constant on its own will cause an accelerating expansion ($\ddot{S} > 0$). When matter and a cosmological constant are both present, either result may occur depending on which effect is dominant. The first integral of equations (5, 7) when $\dot{S} \neq 0$ is the *Friedmann equation*

(9) $$\frac{\dot{S}^2}{S^2} = \frac{\kappa\mu}{3} + \frac{\Lambda}{3} - \frac{k}{S^2}.$$

This is just the Gauss equation relating the 3-space curvature to the 4-space curvature, showing how matter directly causes a curvature of 3-spaces [Ehlers, 1993; Ellis, 1971a]. Because of the spacetime symmetries, the ten EFE are equivalent to the two equations (7, 9). Models of this kind, that is with a Robertson-Walker ('RW') geometry with metric (2, 3) and dynamics governed by equations (5, 7, 9), are called *Friedmann-Lemaître universes* ('FL' for short). The Friedmann equation (9) controls the expansion of the universe, and the conservation equation (5) controls the density of matter as the universe expands; when $\dot{S} \neq 0$, equation (7) will necessarily hold if (5, 9) are both satisfied.

Given a determinate matter description (specifying the equation of state $w = w(\mu, T)$ explicitly or implicitly) for each matter component, the existence and uniqueness of solutions follows both for a single matter component and for a combination of different kinds of matter, for example $\mu = \mu_{bar} + \mu_{rad} + \mu_{cdm} + \mu_\nu$ where we include cold dark matter (cdm) and neutrinos (ν). Initial data for such solutions at an arbitrary time t_0 (eg. today) consists of,

- The *Hubble constant* $H_0 := (\dot{S}/S)_0 = 100h$ km/sec/Mpc;
- A dimensionless *density parameter* $\Omega_{i0} := \kappa\mu_{i0}/3H_0^2$ for each type of matter present (labelled by i);
- If $\Lambda \neq 0$, either $\Omega_{\Lambda 0} := \Lambda/3H_0^2$, or the dimensionless *deceleration parameter* $q_0 := -(\ddot{S}/S)_0 H_0^{-2}$.

Given the equations of state for the matter, this data then determines a unique solution $\{S(t), \mu(t)\}$, i.e. a unique corresponding universe history. The total matter density is the sum of the terms Ω_{i0} for each type of matter present, for example

(10) $\Omega_{m0} = \Omega_{bar0} + \Omega_{rad0} + \Omega_{cdm0} + \Omega_{\nu 0},$

[1] A cosmological constant can also be regarded as a fluid with pressure p related to the energy density μ by $\{p = -\mu c^2 \Leftrightarrow w = -1\}$. For the history of the cosmological constant, see [Earman, 2001; Earman, 2003].

and the total density parameter Ω_0 is the sum of that for matter and for the cosmological constant:

(11) $\Omega_0 = \Omega_{m0} + \Omega_{\Lambda 0}$.

Evaluating the Raychaudhuri equation (7) at the present time gives an important relation between these parameters: when the pressure term p/c^2 can be ignored relative to the matter term μ (as is plausible at the present time),[2]

(12) $q_0 = \frac{1}{2}\,\Omega_{m0} - \Omega_{\Lambda 0}$.

This shows that a cosmological constant Λ can cause an acceleration (negative q_0); if it vanishes, the expression simplifies: $\Lambda = 0 \Rightarrow q = \frac{1}{2}\,\Omega_{m0}$, showing how matter causes a deceleration of the universe. Evaluating the Friedmann equation (9) at the time t_0, the spatial curvature is

(13) $K_0 := k/S_0^2 = H_0^2\,(\Omega_0 - 1)$.

The value $\Omega_0 = 1$ corresponds to spatially flat universes ($K_0 = 0$), separating models with positive spatial curvature ($\Omega_0 > 1 \Leftrightarrow K_0 > 0$) from those with negative spatial curvature ($\Omega_0 < 1 \Leftrightarrow K_0 < 0$).

The FL models are the standard models of modern cosmology, surprisingly effective in view of their extreme geometrical simplicity. One of their great strengths is their explanatory role in terms of making explicit the way the local gravitational effect of matter and radiation determines the evolution of the universe as a whole, this in turn forming the dynamic background for local physics (including the evolution of the matter and radiation).

2.1.1 The basic solutions

For baryons (pressure-free matter) and non-interacting radiation, the Friedmann equation (9) takes the form

(14) $$\frac{3\dot{S}^2}{S^2} = \frac{A}{S^3} + \frac{B}{S^4} + \frac{\Lambda}{3} - \frac{3k}{S^2}$$

where $A := \kappa\mu_{bar0}S_0^3$ and $B := \kappa\mu_{rad0}S_0^4$. The behaviour depends on the cosmological constant Λ [Robertson, 1933; Rindler, 2001].

When $\Lambda = 0$, the universe starts off at a very dense initial state — according to the classical theory, an initial singularity where the density and curvature go infinite (see Sec. 2.1.2). Its future fate depends on the value of the spatial curvature, or equivalently the density parameter Ω_0. The universe expands forever if $\{k = 0 \Leftrightarrow \Omega_0 = 1\}$ or $\{k < 0 \Leftrightarrow \Omega_0 < 1\}$, but collapses to a future singularity if $\{k > 0 \Leftrightarrow \Omega_0 > 1\}$. Thus $\Omega_0 = 1$ *corresponds to the critical density* μ_{crit} *separating* $\Lambda = 0$ *FL models that recollapse in the future from those that expand forever*, and Ω_0 is just the ratio of the matter density to this critical density:

[2]Assuming we represent 'dark energy' (Sec. 2.3.6) as a cosmological constant.

(15) $\{\Omega_{crit} = 1 \Leftrightarrow \kappa\mu_{crit} = 3H_0^2\} \Rightarrow \Omega_0 := \kappa\mu_0/3H_0^2 = \mu_0/\mu_{crit}$.

When $\Lambda < 0$, all solutions start at a singularity and recollapse.

When $\Lambda > 0$, if $k = 0$ or $k = -1$ all solutions start at a singularity and expand forever. If $k = +1$ there can again be models with a singular start, either expanding forever or collapsing to a future singularity. However in this case a static solution (the Einstein static universe) is also possible, as well as models asymptotic to this static state in either the future or the past. Furthermore models with $k = +1$ can bounce (collapsing from infinity to a minimum radius and re-expanding).

The dynamical behaviour of these models has been investigated in depth: first for dust plus a cosmological constant [Robertson, 1933; Rindler, 2001], followed by perfect fluids, fluids with bulk viscosity, kinetic theory solutions, and scalar field solutions. Current models employ a realistic mixture of matter components (baryons, radiation, neutrinos, cold dark matter, a scalar field, and perhaps a cosmological constant). Informative phase planes show clearly the way higher symmetry (self-similar) models act as attractors and saddle points for the other models [Madsen and Ellis, 1988; Ehlers and Rindler, 1989].

The simplest expanding solutions are the following:

1. The *Einstein-de Sitter* model, for which $\{p = 0, \Lambda = 0, k = 0\} \Rightarrow \Omega_0 = 1$. This is the simplest expanding non-empty solution:

   (16) $S(t) = C\,t^{2/3}$

   starting from a singular state at time $t = 0$ (C is an arbitrary constant). Its age (the proper time since the start of the universe) when the Hubble constant takes the value H_0 is $\tau_0 = \frac{2}{3H_0}$. This is a good model of the expansion of the universe since radiation domination ended until the recent times when a cosmological constant started to dominate the expansion. It is also a good model of the far future universe if $k = 0$ and $\Lambda = 0$.

2. The *Milne* model, for which $\{\mu = p = 0, \Lambda = 0, k = -1\} \Rightarrow \Omega_0 = 0$, giving a linearly expanding empty solution:

   (17) $S(t) = C\,t.$

   This is just flat space-time as seen by a uniformly expanding set of observers [Rindler, 2001, pp. 360-363], singular at $t = 0$. Its age is $\tau_0 = \frac{1}{H_0}$. It is a good model of the far future universe if $k < 0$ and $\Lambda = 0$.

3. The *de Sitter* universe, for which $\{\mu = p = 0, \Lambda \neq 0, k = 0\} \Rightarrow \Omega_0 = 0$, giving the steady state expanding empty solution:[3]

   (18) $S(t) = C\,\exp(Ht)$,

[3]The Steady State universe of Bondi, Hold and Hoyle [Bondi, 1960] utilised this metric, but was non-empty as they abandoned the EFE.

where C and H are constants. As the expansion rate is constant forever, there is no start and its age is infinite.[4] It is a good model of the far future universe for those cases which expand forever with $\Lambda > 0$. It can alternatively be understood as a solution with $\Lambda = 0$ and containing matter with the exceptional equation of state $\mu + p/c^2 = 0$. There are other RW forms of the de Sitter Universe: a geodesically complete form with $k = +1$, $S(t) = S_0 \cosh Ht$ (a regular bounce), and another geodesically incomplete form with $k = -1$, $S(t) = S_0 \sinh Ht$ (a singular start). This lack of uniqueness is possible because *this is a spacetime of constant curvature*, with no preferred timelike directions or space sections [Schrödinger, 1956; Hawking and Ellis, 1973; Rindler, 2001].[5]

2.1.2 An initial singularity?

The above are specific models: what can one say generically? When the inequality (8) is satisfied, one obtains directly from the Raychaudhuri equation (7) the

> **Friedmann-Lemaître Universe Singularity Theorem** [Ehlers, 1993; Ellis, 1971a]: In a FL universe with $\Lambda \leq 0$ and $\mu + 3p/c^2 > 0$ at all times, at any instant t_0 when $H_0 \equiv (\dot{S}/S)_0 > 0$ there is a finite time t_*: $t_0 - (1/H_0) < t_* < t_0$, such that $S(t) \to 0$ as $t \to t_*$; the universe starts at a space-time singularity there, with $\mu \to \infty$ and $T \to \infty$ if $\mu + p/c^2 > 0$.

This is not merely a start to matter — it is a start to space, to time, to physics itself. It is the most dramatic event in the history of the universe: it is the start of existence of everything. The underlying physical feature is the non-linear nature of the EFE: going back into the past, the more the universe contracts, the higher the active gravitational density, causing it to contract even more. The pressure p that one might have hoped would help stave off the collapse makes it even worse because (consequent on the form of the EFE) p enters algebraically into the Raychaudhuri equation (7) with the same sign as the energy density μ. Note that the Hubble constant gives an estimate of the age of the universe: the time $\tau_0 = t_0 - t_*$ since the start of the universe is less than $1/H_0$.

This conclusion can in principle be avoided by a cosmological constant, but in practice this cannot work because we know the universe has expanded by at least a ratio of 6, as we have seen objects at a redshift[6] of 5; from (14), the cosmological constant would have to have an effective magnitude at least $6^3 = 216$ times the present matter density to dominate and cause a turn-around then or at any earlier time, and so would be much bigger than its observed present upper limit (of the

[4]It is however singular in that it is geodesically incomplete; this metric covers only half the de Sitter hyperboloid [Schrödinger, 1956; Hawking and Ellis, 1973].

[5]There is also a static (non-RW) form of the metric — the first form of the metric discovered.

[6]The redshift z for light emitted at t_e and observed at t_0 is related to the expansion by $1 + z = S(t_0)/S(t_e)$, see Sec. 2.3.3.

same order as the present matter density). Accordingly, no turn around is possible while classical physics holds [Ehlers and Rindler, 1989]. However energy-violating matter components such as a scalar field (Sec. 2.6) can avoid this conclusion, if they dominate at early enough times; but this can only be when quantum fields are significant, when the universe was at least 10^{12} smaller than at present.

Because $T_{rad} \propto S^{-1}$ (eqn.(6)), a major conclusion is that *a* Hot Big Bang *must have occurred; densities and temperatures must have risen at least to high enough energies that quantum fields were significant*, at something like the GUT energy. The universe must have reached those extreme temperatures and energies at which classical theory breaks down.

2.2 The hot big bang

The matter and radiation in the universe gets hotter and hotter as we go back in time towards the initial quantum state, because it was compressed into a smaller volume. In this *Hot Big Bang* epoch in the early universe, we can use standard physical laws to examine the processes going on in the expanding mixture of matter and radiation [Weinberg, 1972; Perkins, 2005]. A key feature is that about 300,000 years after the start of the Hot Big Bang epoch, nuclei and electrons combined to form atoms. At earlier times when the temperature was higher, atoms could not exist, as the radiation then had so much energy it disrupted any atoms that tried to form into their constituent parts (nuclei and electrons). Thus at earlier times matter was ionized, consisting of negatively charged electrons moving independently of positively charged atomic nuclei. Under these conditions, the free electrons interact strongly with radiation by Thomson scattering. Consequently matter and radiation were tightly coupled in equilibrium at those times, and the Universe was opaque to radiation. When the temperature dropped through the ionization temperature of about 4000K, atoms formed from the nuclei and electrons, and this scattering ceased: the Universe became very transparent (today we are able to see galaxies at enormous distances from us). The time when this transition took place is known as the *time of decoupling* — it was the time when matter and radiation ceased to be tightly coupled to each other, at a redshift $z_{dec} \simeq 1100$ [Dodelson, 2003]. By (6), the universe was radiation dominated ($\mu_{rad} \gg \mu_{mat}$) at early times and matter dominated ($\mu_{rad} \ll \mu_{mat}$) at late times;[7] matter-radiation density equality occurred significantly before decoupling (the temperature T_{eq} when this equality occurred was $T_{eq} \simeq 10^4$K; at that time the scale factor was $S_{eq} \simeq 10^4 S_0$, where S_0 is the present-day value). The dynamics of both the background model and of perturbations about that model differ significantly before and after S_{eq} [Dodelson, 2003].

[7]The dynamically dominant Cold Dark Matter (Sec. 2.3.6) obeys the same density law (6) as baryons.

2.2.1 Cosmic Blackbody Radiation

Radiation was emitted by matter at the time of decoupling, thereafter travelling freely to us through the intervening space. When it was emitted, it had the form of blackbody radiation, because this is a consequence of matter and radiation being in thermodynamic equilibrium at earlier times. Thus *the matter at $z = z_{dec}$ forms the Last Scattering Surface* (LSS) *in the early universe, emitting Cosmic Blackbody Background Radiation*[8] ('CBR') *at 4000K, that since then has travelled freely with its temperature T scaling inversely with the scale function of the universe.*[9] As the radiation travelled towards us, the universe expanded by a factor of about 1100; consequently by the time it reaches us, it has cooled to 2.75 K (that is, about 3 degrees above absolute zero, with a spectrum peaking in the microwave region), and so is extremely hard to observe. It was however detected in 1965, and its spectrum has since been intensively investigated, its blackbody nature being confirmed to high accuracy [Partridge, 1995]. Its existence is now taken as solid proof both that the Universe has indeed expanded from a hot early phase, and that standard physics applied unchanged at that era in the early universe.

2.2.2 Particle interactions and element formation

The thermal capacity of the radiation is hugely greater than that of the matter. At very early times before decoupling, the temperatures of the matter and radiation were the same (because they were in equilibrium with each other), scaling as $1/S(t)$ (eqn.(6)). The early universe exceeded any temperature that can ever be attained on Earth or even in the centre of the Sun; as it dropped towards its present value of 3 K, successive physical reactions took place that determined the nature of the matter we see around us today. At very early times and high temperatures, only elementary particles can survive and even neutrinos had a very small mean free path; as the universe cooled down, neutrinos decoupled from the matter and streamed freely through space. At these times the expansion of the universe was radiation dominated, and we can approximate the universe then by models with $\{k = 0,\ w = 1/3,\ \Lambda = 0\}$, the resulting simple solution of (14) uniquely relating time to temperature:

$$(19)\quad S(t) = S_0\, t^{1/2},\ \ t = 1.92 \sec \left[\frac{T}{10^{10} K}\right]^{-2}.$$

(There are no free constants in the latter equation).

At very early times, even neutrinos were tightly coupled and in equilibrium with the radiation; they decoupled at about 10^{10}K [Dodelson, 2003, pp. 44-46], resulting in a relic neutrino background density in the universe today of about $\Omega_{\nu 0} \simeq 10^{-5}$ if they are massless (but it could be higher depending on their masses). Key events in the early universe are associated with out of equilibrium phenomena

[8]This is often called "Cosmic Microwave Background", or CMB for short. However it is only microwave at the present epoch.

[9]This scaling for freely propagating radiation follows from the discussion in Sec. 2.3.3.

[Dodelson, 2003, p. 58]. An important event was the era of *nucleosynthesis*, the time when the light elements were formed. Above about 10^9K, nuclei could not exist because the radiation was so energetic that as fast as they formed, they were disrupted into their constituent parts (protons and neutrons). However below this temperature, if particles collided with each other with sufficient energy for nuclear reactions to take place, the resultant nuclei remained intact (the radiation being less energetic than their binding energy and hence unable to disrupt them). Thus the nuclei of the light elements — deuterium, tritium, helium, and lithium — were created by neutron capture. This process ceased when the temperature dropped below about 10^8 K (the nuclear reaction threshold). In this way, the proportions of these light elements at the end of nucleosynthesis were determined; they have remained virtually unchanged since. The rate of reaction was extremely high; all this took place within the first three minutes of the expansion of the Universe. One of the major triumphs of Big Bang theory is that *theory and observation are in excellent agreement provided the density of baryons is low:* $\Omega_{bar0} \simeq 0.044$. *Then the predicted abundances of these elements* (25% *Helium by weight,* 75% *Hydrogen, the others being less than* 1%) *agrees very closely with the observed abundances.* Thus the standard model explains the origin of the light elements in terms of known nuclear reactions taking place in the early Universe [Schramm and Turner, 1998]. However heavier elements cannot form in the time available (about 3 minutes).

In a similar way, physical processes in the very early Universe (before nucleosynthesis) can be invoked to explain the ratio of matter to anti-matter in the present-day Universe: a small excess of matter over anti-matter must be created then in the process of *baryosynthesis*, without which we could not exist today (if there were no such excess, matter and antimatter would have all annihilated to give just radiation [Silk, 2005]). However other quantities (such as electric charge) are believed to have been conserved even in the extreme conditions of the early Universe, so their present values result from given initial conditions at the origin of the Universe, rather than from physical processes taking place as it evolved. In the case of electric charge, the total conserved quantity appears to be zero: after quarks form protons and neutrons at the time of baryosynthesis, there are equal numbers of positively charged protons and negatively charged electrons, so that at the time of decoupling there were just enough electrons to combine with the nuclei and form uncharged atoms (it seems there is no net electrical charge on astronomical bodies such as our galaxy; were this not true, electromagnetic forces would dominate cosmology, rather than gravity).

After decoupling, matter formed large scale structures through gravitational instability [Bothun, 1998, pp. 183-222] which eventually led to the formation of the first generation of stars [Silk, 2005] and is probably associated with the re-ionization of matter [Dodelson, 2003, p. 73]. However at that time planets could not form for a very important reason: there were no heavy elements present in the Universe. The first stars aggregated matter together by gravitational attraction, the matter heating up as it became more and more concentrated, until its temperature exceeded the thermonuclear ignition point and nuclear reactions started burning hydrogen to form helium. Eventually more complex nuclear reactions started in concentric spheres around the centre, leading to a build-up of heavy elements (carbon, nitrogen, oxygen for example), up to iron. These elements can form in stars because there is a long time available (millions of years) for the reactions to take place. Massive stars burn relatively rapidly, and eventually run out of nuclear fuel. The star becomes unstable, and its core rapidly collapses because of gravitational attraction. The consequent rise in temperature blows it apart in a giant explosion, during which time new reactions take place that generate elements heavier than iron; this explosion is seen by us as a Supernova ("New Star") suddenly blazing in the sky, where previously there was just an ordinary star. Such explosions blow into space the heavy elements that had been accumulating in the star's interior, forming vast filaments of dust around the remnant of the star. It is this material that can later be accumulated, during the formation of second generation stars, to form planetary systems around those stars. Thus *the elements of which we are made* (*the carbon, nitrogen, oxygen and iron nuclei for example*) *were created in the extreme heat of stellar interiors, and made available for our use by supernova explosions*. Without these explosions, we could not exist.

2.3 Cosmological Observations

Cosmological models only become meaningful when related to astronomical observations [Hoyle, 1960; Sandage, 1961; Ellis, 1971a; Weinberg, 1972]. These are of two kinds: astronomical observations of distant matter tells us what was happening far away in the universe and (because of the finite speed of light) a long time ago. On the other hand observations of nearby objects (matter on Earth, the solar system, nearby stars for example) when related to theories of origins tell us what was happening very near our past world line a very long time ago. The first set of observations may be characterized as "null cone" observations, the second as "geological" observations, one of the most important being the determination of local element abundances, which are then related to nucleosynthesis calculations (Sec. 2.2.2).

Observations are totally dependent on telescope and detector technology [Harwit, 1984; Bothun, 1998]. After the initial establishment of distance scales and the basic evidence of cosmic homogeneity and expansion in the 1920s and 1930s, progress was slow until the 1960s when observations were extended from the optical to the entire electromagnetic spectrum. In recent decades cosmology has changed from a data-poor to a data-rich subject. Massive new data sets are now available because of the extraordinary improvement of telescope, detector, and computer technology in recent decades, particularly the advent of new detectors such as Charge Coupled Devices (CCD's) and fibre optics (enabling simultaneous measurement of hundreds of redshifts). We now have not only optical, ultraviolet, and infrared observations of galaxies, determining luminosities and spectra with unprecedented sensitivity, but also radio, X-ray, and gamma-ray sky surveys. Galaxies have been detected up to a redshift of 6 and we have identified many quasi-stellar objects and gamma-ray bursters as well as multiple images of very distant gravitationally-lensed galaxies [Harwit, 1984]. Large-scale structures (clusters of galaxies, superclusters, walls, and voids) have been identified, with associated large-scale velocity flows [Bothun, 1998, pp. 85-137].

In addition to large-scale number-count and redshift surveys, we have measured the background radiation spectrum and anisotropies at all wavelengths. We identify the radiation as 'background' precisely when it is constant on very small angular scales (as opposed to discrete sources, which appear as isolated objects). There is a complex relation of this radiation to the intergalactic matter density and thermal history. The most important component of the background radiation is the Cosmic Blackbody Radiation ('CBR') mentioned above (Sec. 2.2); detailed observations have mapped its temperature over the whole sky at a sensitivity of better than one part in 10^5. However other components of the background radiation (X-ray and radio in particular) convey important information on the temperature and density of intergalactic matter, and hence strongly restrict its possible thermal history. For example hot matter emits X-rays, so the X-ray background measurement restricts the amount of hot intergalactic matter allowed; while neutral hydrogen strongly absorbs Ultra-Violet radiation to give the Lyman alpha spectral absorption line, so absence of such absorption gives strong limits on the amount of neutral hydrogen and hence on the temperature of intergalactic matter.

2.3.1 Isotropy

The first important point about cosmological observations is that *when averaged on a large enough physical scale* (*clusters of galaxies and above*) *they are statistically isotropic about us*; there is no direction apparently pointing to the centre of the universe. The *high degree of isotropy of the CBR* strongly supports this conclusion: its temperature is the same in all directions about us to better than one part in 10,000 after we have allowed for the motion of the Earth relative to the cosmos (about 250 km/sec), which creates a temperature dipole at one part in a thousand.[10] Any inhomogeneities or anisotropies in the matter distribution lead to anisotropies in this radiation, as recently measured at only one part in 10^5 by the extremely sensitive detectors of the COBE and WMAP satellites. This high degree of isotropy is the major reason we believe the Universe itself is spatially homogeneous and isotropic to a good approximation (see Sec. 4.2.2), providing good observational support for use of the FL universe models as accurate models of the observed region of the universe.

2.3.2 Distance scale and ages

The underlying problem in all astronomy is determining the distances of observed objects. This is done by a 'cosmic distance ladder' [Bothun, 1998, pp. 25-83]

[10]The CBR dipole that could be interpreted as due to a major cosmological inhomogeneity is rather interpreted as being due to our motion ('peculiar velocity') relative to a spatially homogeneous universe.

whereby nearest objects have their distance determined by parallax (i.e. essentially by local trigonometry); and more distant ones by a series of consecutive distance indicators (Cepheid variables, RR Lyrae variables, brightest red supergiants) until at a cosmological distance, redshift z is a primary distance indicator, but is contaminated by local velocities of matter relative to the rest-frame of the universe. Other distance indicators (for example the Tully-Fisher method, the luminosity function of planetary nebulae, the globular cluster luminosity function, surface brightness fluctuations) serve to refine the estimates [Bothun, 1998].

Closely associated with the distance scale is determination of the Hubble constant H_0 (the present rate of expansion of the universe), because estimates of the size of the observable region of the universe scale with the Hubble constant. But the Hubble constant also determines the age of the universe, so its determination underlies a crucial consistency condition for cosmology: *the age of objects in the universe (rocks, planets, stars, star clusters, galaxies) must be less than the age of the universe.* This condition has been a cause of concern ever since we have had good estimates of ages and of the Hubble constant.[11] It seems not to be violated by current observations of low redshift objects given the current estimates of $H_0 \simeq 70$ km/sec/Mpc, giving an age of the universe of about 15 billion years whereas the oldest star clusters seem to be about 14 billion years old. However it is very tight, perhaps even problematic, for very distant (and so much younger) objects [Jain and Dev, 2005].

2.3.3 Observational relations

Light travels on null geodesics $x^a(\lambda)$ in spacetime (the tangent vector $k^a := dx^a/d\lambda$ is such that $k^a{}_{;b}k^b = 0$, $k^a k_a = 0$). In a RW geometry, it suffices to consider only radial null geodesics (by the symmetries of the model, these are equivalent to generic geodesics). Then from (2) we find that for light emitted at time t_e and received at time t_0, the comoving radial distance $u(t_0, t_1) := r_0 - r_1$ between comoving emitters and receivers is given by

$$(20)\quad \{ds^2 = 0,\ d\theta = 0 = d\phi\} \;\Rightarrow\; u(t_0, t_1) = \int_{t_1}^{t_0} \frac{dt}{S(t)} = \int_{S_1}^{S_0} \frac{dS}{S\dot{S}}$$

with $\dot{S}$ given by the Friedmann equation (9). The key quantities related to cosmological observations are redshift, area distance (or apparent size), and local volume corresponding to some increment in distance (determining number counts) [Sandage, 1961; Ellis, 1971a; Weinberg, 1972]. The redshift z measured for an object emitting light of wavelength λ_e that is observed with wavelength λ_0 is given by

[11]Indeed Hubble himself never fully accepted the expanding universe theory because of age difficulties, preferring to refer to a redshift-distance relation rather than a velocity distance relation [Hubble, 1936]. However the problem has been eased by a series of revisions of the value of the Hubble constant since then, due to a better understanding of the primary distance indicators.

(21) $1 + z := \frac{\lambda_0}{\lambda_e} = (1 + z_c)(1 + z_v)$,

where z_v is the redshift caused by the local peculiar motion of the object observed ($z_v = 0$ for comoving objects), and z_c is the cosmological redshift given by

(22) $1 + z_c = \frac{S(t_0)}{S(t_e)}$.

From eqn.(21), the same ratio of observed to emitted light holds for all wavelengths: a key identifying property of redshift. The problem in using redshifts of objects as a distance indicator is to separate out the cosmological from the Doppler components, which lead to redshift-space distortions [Dodelson, 2003, pp. 275-282]; this can reasonably be done for a cluster of galaxies by appropriate averaging over cluster members ($\langle z_v \rangle = 0$ for a comoving cluster). The area distance r_0 of an object at redshift z_c and of linear size l which subtends angular size α is given by[12]

(23) $r_0(z_c) := \frac{l}{\alpha} = f(u)S(t_e)$.

Thus measures of apparent sizes will determine the area distance if the source physical size is known. The flux of radiation F measured from a point source of luminosity L emitting radiation isotropically is given by the fraction of radiant energy emitted by the source in a unit of time that is received by the telescope:

(24) $F = \frac{L}{4\pi} \frac{1}{f^2(u)S^2(t_0)(1+z)^2}$;

(the two redshift factors account firstly for time dilation observed between observer and source, and secondly for loss of energy due to redshifting of photons). The source's *apparent magnitude* m is defined from the flux: $m = -2.5 log_{10} F + const.$ On using (22, 23), equation (24) becomes

(25) $F = \frac{L}{4\pi} \frac{1}{r_0^2 (1+z)^4}$.

showing that measures of magnitudes will determine the area distance if the source's intrinsic luminosity is known. On using (23) it follows from (25) that *the point-wise surface brightness of extended objects (the flux received per unit solid angle) depends only on redshift* [Kristian and Sachs, 1966; Ellis, 1971a] — a key feature in determining detection probabilities and in gravitational lensing observations. It further follows from this result that *a blackbody spectrum emitted at temperature T_e when observed with a redshift z remains a blackbody spectrum but with observed temperature $T_0 = T_e/(1 + z)$* — a crucial feature in analyzing the CBR observations.

[12]This depends only on z_c because apparent shapes and sizes are independent of the motion of the source.

Using the Friedmann equation and the relevant equation of state for matter, the area distance can be determined as a function of redshift z_c in terms of the Hubble constant H_0, deceleration parameter q_0, and cosmological constant Λ. In the case of pressure-free matter with vanishing cosmological constant, one obtains from (20), (9), (22), and (23)[13] the Mattig relation [Sandage, 1961]

$$(26) \quad r_0(z_c) = \frac{1}{H_0 q_0^2} \frac{(q_0 - 1)(1 + 2q_0 z_c)^{1/2} + (q_0(z_c - 1) + 1)}{(1 + z_c)^2} \,.$$

Consequently measures of either apparent size of sources of known physical size, or of radiant flux from sources of known intrinsic luminosity, will determine the deceleration parameter q_0. Generalizations of this relation hold if a cosmological constant or radiation is present. An interesting aspect is that there is a minimum apparent size for objects of fixed physical size at some redshift $z_c = z_*$ depending on the density parameter and the cosmological constant. The past light cone of the observer attains a maximum area at z_*; the entire universe acts as a gravitational lens for further off objects, magnifying their apparent size so that very distant objects can appear to have the same angular size as nearby ones [Hoyle, 1960; Sandage, 1961; Ellis, 1971a]. For the Einstein-de Sitter universe, the minimum angular diameter is at $z_* = 1.25$; in low density universes, it occurs at higher redshifts.

The number of objects seen in a solid angle $d\Omega$ for a distance increment du (characterized by increments dz, dm in the observable variables z and m) is given by

$$(27) \quad dN = W(u)\, \rho(t_e) S^3(t_e) f(u) du d\Omega$$

where the detection probability or 'selection function' is $W(u)$ [Dodelson, 2003, p. 263] and $\rho(t_e)$ is the number density of objects at the time of emission (spatial homogeneity is expressed in the fact that this is independent of the spatial coordinates). The observed total number N of objects measured in a survey is given by integrating from the observer to the survey limit: in terms of the radial coordinate r_e of the source (which can be related to redshifts or magnitudes), $N = \int_{r_0}^{r_e} dN$. If the number of objects is conserved (e.g. observing galaxies in an epoch when they are neither created nor destroyed), $\rho(t_e) = \rho(t_0)(1+z)^3$ and we find from (27) that in the idealized case when W is independent of distance (a reasonable assumption for relatively nearby objects),

$$(28) \quad N = W \rho(t_0) d\Omega \int_{r_0}^{r_e} f(u) du.$$

The simple integral has to be separately done for the cases $k = +1, 0, -1$ [Sandage, 1961].

The above equations enable one to determine observational relations between observable variables, for example (m, z), (α, z) or (N, m) relations for objects with known intrinsic properties (known size or luminosity, for example), and so

[13] Or, more elegantly, from the geodesic deviation equation (see [Ellis and van Elst, 1999b]).

to observationally determine q_0. These relations have to be modified if there is absorption by an intergalactic medium, gravitational lensing, or anisotropic emission of radiation; and detailed comparisons with observations have to take into account the spectrum of the source as well as source detection and identification probabilities [Harwit, 1984]. Here we encounter *the contrast between image and reality*: there can be many objects out there that we either do not detect, or do not recognize for what they are [Disney, 1976]. An "observational map" relating source properties to the nature of their images gives a useful view of how this occurs [Ellis *et al.*, 1984].

One important feature here is that *a specific object will look completely different at different wavelengths* (optical, radio, X-ray for example); indeed it may be detectable at one wavelength and not at another. This shows very clearly how our images of reality are dependent on the detectors we use. To get a full picture of what is out there, we need to use multiple modes of investigation — imaging at all wavelengths together with intensity, spectral, and polarization measurements [Harwit, 1984], as well as watching for time variations. A second important feature is observational selection effects such as the *Malmquist bias* — if we have a population of objects with different luminosities, at large distances we will only see the more luminous objects (the fainter ones will not be detected); hence *the average luminosity will appear to increase with distance*, but this is just an observational effect rather than the real state of affairs. Using different detection thresholds controls this effect to some degree.

2.3.4 Number Counts and the visible matter density

Number counts of galaxies as a function of redshift or luminosity show approximate spatial homogeneity of the universe [Hubble, 1936]. However counts of radio sources and quasi-stellar objects (qso's) show that the universe has not been in a steady state as proposed by Bondi, Gold, and Hoyle [Bondi, 1960]. Indeed *number counts are only compatible with a RW geometry if there has been evolution of source numbers and/or luminosities* [Sciama, 1971].

Number counts also give estimates of the density of visible (luminous) matter in the universe: $\Omega_{vm0} \simeq 0.015$. This is very low relative to the critical density ($\Omega_0 = 1$) and is also considerably less than the amount of baryons determined by nucleosynthesis studies ($\Omega_{bar0} \simeq 0.044$, see Sec. 2.2.2). Thus *much of the baryonic matter in the universe is in some hidden (non-luminous) form* [Bothun, 1998, pp. 223-272], e.g. burnt out stars [Hogan, 1999].

2.3.5 Apparent Luminosities and sizes: Dark Energy

Apparent sizes or luminosities as a function of redshift can be used to determine the deceleration parameter q_0 (Sec. 2.1) if the intrinsic source sizes or luminosities are known. The problem is that until recently there were not known enough galaxies or other objects of standard size or luminosity to use to determine q_0, and scatter in their properties leads to biassing of observations by the Malmquist effect. However this dramatically changed with recent observations of the decay curves of the luminosity of supernovae in distant galaxies. It turns out that the peak luminosity of type Ia supernovae is closely correlated with their light curve decay time, for the first time giving reliable 'standard candles' for galaxies at large distances [Perlmutter *et al.*, 1998]. The conclusion from these observations is that, rather than slowing down as expected, *the rate of expansion of the universe is speeding up at a rate corresponding to a cosmological constant with* $\Omega_{\Lambda 0} = 0.7$. This evidence is concordant with that from CBR observations and number counts [Dodelson, 2003; Silk, 2005].

The nature of the field or matter causing this acceleration is unclear. Its equation of state $w := pc^2/\mu$, is unknown, and many physical and unphysical proposals are being made in this regard. From (7), it has to violate the strong energy condition (8) and so must have a large negative pressure. It could indeed be due to a cosmological constant ($w = -1$), which would have dominated the expansion of the universe since a redshift $z \simeq 0.33$, and would have been negligible earlier (and is also negligible on small scales — it does not affect local astrophysics). However it could also be some other form of matter or field with effective negative pressure so that $w < -1/3$, as can happen in the case of a scalar field (see eqn.(33) below). In that case it is called 'quintessence'. There are many speculations as to what this might be, but there is no clarity on the matter. One should note here that alternative explanations of the observations are possible, for they can be exactly reproduced by a spherically symmetric inhomogeneous universe model where we are near the centre [Mustapha *et al.*, 1998], or could at least partly be due to the backreaction of inhomogeneities on the cosmic expansion [Ellis and Buchert, 2006] or the effect of inhomogeneities on the effective area distance [Kantowski *et al.*, 1995; Kantowski, 1998]. These alternatives are being investigated, but the most probable cause remains some unknown kind of matter or field with effective negative energies.

In summary, the standard gravitational equations together with the supernovae observations imply *presence of a cosmological constant or some equivalent form of 'dark energy' with a large effective negative energy density* μ_{grav} *(due to negative pressure) dominating the present expansion of the universe; its physical nature is unknown.* There is no known physics reason why this force should exist at this level where it is just detectable — quantum field theory relates the cosmological constant to the zero-point energy of the vacuum, and suggests it should be enormously larger than observed [Weinberg, 1989; 2000a; 2000b; Rugh and Zinkernagel, 2002; Zinkernagel, 2002; Susskind, 2005]. It is a major mystery why it exists at the small (just detectable) level that observations indicate [Seife, 2003]. A key aspect of present day cosmology is trying on the one hand to observationally determine the effective equation of state of this 'dark energy' (running the field equations backwards to obtain $w(z)$ from the observations [Saini *et al.*, 2000], and in particular determining whether w is constant or varying over time), and on the other attempting to give a plausible theoretical explanation for its physical origin.

2.3.6 Matter Distribution and Motion: Dark Matter

Detailed studies have been made of the distribution of galaxies and their motions. They occur in clusters, in turn making up superclusters imbedded in vast walls surrounding relatively sparsely populated intergalactic voids. The galaxy *luminosity function* characterizes the numbers of galaxies occurring within each luminosity class; the *covariance function* characterizes their spatial clustering [Peebles, 1993b; Dodelson, 2003]. Large scale motions occur for galaxies in clusters, and for the clusters themselves. It is easy to conceive of matter that is hard to detect (for example, small rocks distributed through space); studies of galactic rotation curves and of motions of galaxies in clusters [Bothun, 1998, pp. 139-181] imply *the existence of huge amounts of unseen dark matter, dominating the dynamics of the Universe*: its density is $\Omega_{dm0} \simeq 0.3$, much greater than both visible matter ($\Omega_{vm0} = 0.015$) and baryons ($\Omega_{bar0} = 0.044$), but significantly less than the critical density $\Omega_0 = 1$. Thus *the dark matter is non-baryonic*, meaning it has some kind of exotic nature rather than being the protons and neutrons that are the substance of ordinary matter [Seife, 2003]. In contrast to the 'dark energy' discussed in the previous section, dark matter is dynamically effective on astrophysical scales as well as on cosmological scales. Many attempts have been made to identify its nature, for example it might be axions, supersymmetric partners of known particles, quark nuggets, or massive neutrinos [Gribbin and Rees, 1991; Perkins, 2005], but what it is composed of is still unknown. Laboratory searches are under way to try to detect this matter, so far without success. A key question is whether its apparent existence is due to our using the wrong theory of gravity on these scales. This is under investigation, with various proposals for modified forms of the gravitational equations that might explain the observations without the presence of large quantities of dark matter. This is a theoretical possibility, but the current consensus is that this dark matter does indeed exist.

An important distinction is whether dark matter consists of

(i) weakly interacting massive particles that cooled down quickly, thereafter forming *cold dark matter* ('CDM') moving slowly at the time of structure formation (and resulting in a bottom-up process with large scale structure forming from smaller scale structures), or

(ii) particles that have a low mass and cooled slowly, therefore for a long time forming *hot dark matter*, moving very fast at the time of structure formation (and resulting in a top-down galaxy formation scenario).

Structure formation studies currently support the CDM hypothesis, with hierarchical formation of gravitationally bound objects taking place in a complex bottom up process involving interactions of CDM, baryons, and radiation, with dwarf galaxies forming initially [Silk, 2005; Mouri and Taniguchi, 2005] and then aggregating to form larger structures. These studies are based on massive numerical simulations, with initial conditions provided by the inflationary scenario discussed below, see

Sec. 2.6. Unlike 'dark energy', CDM has an ordinary baryonic equation of state (it is a perfect fluid (4) with $p_{cdm} = 0 \Leftrightarrow w_{cdm} = 0$).

Another way of detecting dark matter in clusters is by its gravitational lensing effects [Schneider *et al.*, 1992]. The bending of light by massive objects was one of the classic predictions of General Relativity theory. Rich clusters of galaxies or galaxy cores can cause strong lensing of more distant objects, where multiple images of distance galaxies and qso's occur, sometimes forming rings or arcs; and weaker lensing by closer masses results in characteristic patterns of distortions of images of distant objects. Analysis of multiple images can be used to reconstruct the lensing mass distributions, and statistical analysis of weak lensing patterns of image distortions are now giving us detailed information on the matter distribution in distant galaxies and clusters. These studies show that to get enough lenses in an almost flat cosmology ($\Omega_0 \simeq 1$) requires the presence of a cosmological constant — there cannot be a critical density of dark matter present [Dodelson, 2003; Silk, 2005].

A key feature of present-day cosmology is attempts to identify the nature of this dark matter, and if possible to detect it in a laboratory situation. While observations favour the CDM scenario, some residual problems as regards the emergence of fine-scale structure still need resolution [Silk, 2005].

2.3.7 The CBR Power spectrum

The CBR angular anisotropies are characterized by an angular power spectrum showing the amount of power in perturbations at each physical scale on the LSS [Bennet et al., 2003; Seife, 2003; Dodelson, 2003]. In the time from the the end of inflation to the LSS, modes of different wavelengths can complete a different number of oscillations. This translates the characteristic periods of the waves into characteristic lengths on the LSS, leading to a series of maxima ('acoustic peaks') and minima in the inhomogeneities on the LSS and consequently in the CBR angular anisotropy power spectrum [Hu and Sugiyama, 1995b; Hu and Sugiyama, 1995a; Peacock, 1999; Perkins, 2005]. These inhomogeneities then form the seeds for structure formation and so are related to the power spectrum of physical scales for structures that form later. They are characterised by a (3-dimensional) spatial power spectrum on the LSS; because we receive the observed CBR radiation from the 2-sphere $S_{2:LSS}$ where our past light cone intersects the LSS, this is seen by us as a 2-dimensional power spectrum of anisotropies on the sky (characterised by the unit sphere S_2 of all direction vectors e_a: $e^a e_a = 1$, $e^a u_a = 0$).

The apparent angular size of the largest CBR peak (about 1^o) allows estimates of the area distance to the LSS and hence of the density of matter in the universe for various values of the cosmological constant, and determines the major cosmological parameters [Spergel *et al.*, 2003]:

> *"By combining WMAP data with other astronomical data sets, we constrain the geometry of the universe:* $\Omega_{tot} = 1.02 \pm 0.02$*, the equation of state of the dark energy,* $w < -0.78$ *(95% confidence limit), and the*

> *energy density in neutrinos,* $\Omega_\nu h^2 < 0.0076$ *(95% confidence limit). For 3 degenerate neutrino species, this limit implies that their mass is less than 0.23 eV (95% confidence limit). The WMAP detection of early reionization rules out warm dark matter."*

There is however a problem here: while the agreement of theory and observations for all small angular scales is remarkable, there is a divergence at the largest angular scales: the observations show less power than expected. Specifically, the quadrupole and octopole are much lower than theory predicts. Also the axes of the quadrupole and octopole are very precisely aligned with each other, and there are other angular anomalies [Starkman and Schwarz, 2005]. These effects might be due to (i) observational contamination by the galaxy (which gets in the way of our view of the LSS), (ii) a contingent ('chance') event (it represents 'cosmic variance', discussed below, see Sec. 3), (iii) our living in a well-proportioned 'small universe' which is spatially closed so that there is a maximum size to possible fluctuations [Weeks *et al.*, 2003], or (iv) some unexpected new physical effect or deeper problem with our understanding of the early universe. The jury is out as to which the case is; this could turn out to be a crisis for the CBR analysis, but on the other hand one can always just resort to saying it is a statistical fluke (the underlying problem here being the uniqueness of the universe, as discussed in Sec. 3).

There are similar expected peaks in the polarization spectrum of this radiation, and polarization maps should have a mode associated with gravitational waves predicted by inflation to exist in the very early universe (Sec. 2.6); detection of such modes will be a crucial test of inflation [Dodelson, 2003; Sievers *et al.*, 2005b]. Studies of polarization indicate that reionisation of the universe took place as early as a redshift of 17, contrary to what is deduced from qso studies. More detailed studies of anisotropies involve the Sunyaev-Zel'dovich effect (changes in the observed temperature due to scattering by hot matter in galaxy clusters) and gravitational lensing.

There is a huge amount of information in the CBR maps, and their more accurate measurement and interpretation is a central feature of current cosmology [Steinhardt, 1995; Peacock, 1999; Dodelson, 2003; Perkins, 2005].

2.4 Causal and visual horizons

A fundamental feature affecting the formation of structure and our observational situation is the limits arising because causal influences cannot propagate at speeds greater than the speed of light. Thus the region that can causally influence us is bounded by our past null cone. Combined with the finite age of the universe, this leads to the existence of particle horizons limiting the part of the universe with which we can have had causal connection.[14]

A *particle horizon* is by definition comprised by the limiting worldlines of the furthest matter that ever intersects our past null cone [Rindler, 1956; 2001]. This

[14]There are also *event horizons* and *apparent horizons* in some cosmological models [Rindler, 1956; Tipler *et al.*, 1980] and [Rindler, 2001, pp. 376-383].

is the limit of matter that can have had any kind of causal contact with since the start of the universe, characterized by the comoving radial coordinate value

$$(29)\quad u_{ph} = \int_0^{t_0} \frac{dt}{S(t)}\,.$$

The present physical distance to the matter comprising the horizon is

$$(30)\quad d_{ph} = S(t_0)u_{ph}.$$

The key question is whether the integral (29) converges or diverges as we go to the limit of the initial singularity where $S \to 0$. *Horizons will exist in standard FL cosmologies for all ordinary matter and radiation*, for u_{ph} will be finite in those cases; for example in the Einstein-de Sitter universe (see Sec. 2.1.1), $u_{ph} = 3t_0^{1/3}$, $d_{ph} = 3t_0$. We will then have seen only a fraction of what exists, unless we live in a universe with spatially compact sections so small that light has indeed had time to traverse the whole universe since its start; this will not be true for universes with the standard simply-connected topology. Penrose's powerful use of conformal methods (see [Hawking and Ellis, 1973; Tipler *et al.*, 1980]) gives a very clear geometrical picture of the nature of horizons [Ellis and Williams, 2000]. They may not exist in non-FL universes, for example Bianchi (anisotropic) models [Misner, 1969]. In universes with closed spatial sections, a supplementary question arises: Is the scale of closure smaller than the horizon scale? There may be a finite time when causal connectivity is attained, and particle horizons cease to exist. In standard $k = +1$ FL models, this occurs just as the universe reaches the final singularity; if however there is a positive cosmological constant or other effective positive energy density field, it will occur earlier. The horizon always grows, because (29) shows that u_{ph} is a monotonically increasing function of t_0. Despite many contrary statements in the literature, *it is not possible that matter leave the horizon once it has entered.* In a (perturbed) FL model, once causal contact has taken place, it remains until the end of the universe.

The importance of horizons is two-fold: they underlie causal limitations relevant in the origin of structure and uniformity [Misner, 1969; Guth, 1981], and they represent absolute limits on what is testable in the universe [Ellis, 1975; 1980].

2.4.1 Causal limitations

As to causal limitations, horizons are important in regard both to the smoothness of the universe on large scales, and the lumpiness of the universe on small scales. The issue of smoothness is encapsulated in the *horizon problem*: if we measure the temperature of the CBR arriving here from opposite directions in the sky in a standard FL model, it came from regions of the surface of last scattering that can have had no causal contact of any kind with each other since the start of the universe. In a radiation-dominated early universe with scale factor (19), the size of the particle horizon at the time of last scattering appears as an angular scale of about 1^o in the sky today, and corresponds to a comoving physical length of about

400,000 light years when evaluated today. Why then are conditions so similar in these widely separated regions? [Misner, 1968; Guth, 1981; Blau and Guth, 1987; Kolb and Turner, 1990]. Note that this question is of a philosophical rather than physical nature, i.e. there is no contradiction here with any experiment, but rather an unease with an apparent fine tuning in initial conditions. This problem is claimed to be solved by the inflationary universe scenario mentioned below, see Sec. 2.6.

Associated with the existence of horizons is the prediction that physical fields in different regions in the universe should be uncorrelated after symmetry breaking takes place, because they cannot have interacted causally. Consequently, if grand unified theories are correct, topological defects such as monopoles and cosmic strings may be expected as relics of the expansion of the very early universe [Kolb and Turner, 1990]. In a standard cosmology, far too many monopoles are predicted. This is also solved by inflation.

As to the lumpiness, the issue here is that if we believe there was a state of the universe that was very smooth — as indicated at the time of decoupling, by the low degree of anisotropy of the CBR, and represented by the RW geometry of the FL models — then there are limits to the sizes of structures that can have grown since then by causal physical processes, and to the relative velocities of motion that can have been caused by gravitational attraction in the available time (for example, the peculiar motion of our own galaxy relative to the CBR rest frame caused by the huge overdensity called the 'Great Attractor'). If there are larger scale structures or higher velocities, these must have been imprinted in the perturbations at the time of last scattering, for they cannot have been generated in a causal way since that time. They are set into the initial conditions, rather than having arisen by physical causation from a more uniform situation. This is a key factor in the theory of growth of perturbations in the early universe where the expansion damps their growth. The quantity determining the relevant physical scales for local causal influences in an expanding universe is the *comoving Hubble radius* $\lambda_H := (SH)^{-1}$; the way perturbations of wavelength λ develop depends on whether $\lambda > \lambda_H$ or $\lambda < \lambda_H$ [Dodelson, 2003, pp. 146-150].

Actually the domain of causal influence is even more tightly constricted than indicated by the past light cone: the limits coming from the horizon size are limits on what can be influenced by particles and forces acting at the speed of light. However only freely travelling photons, massless neutrinos, and gravitons can move at that speed; and such particles coming from cosmological distances have very little influence on our galaxy or the solar system (indeed we need very delicate experiments to detect them). Any massive particles, or massless particles that are interacting with matter, will travel slower (for example before decoupling, light has a very small mean free path and information will travel only by sound waves and diffusion in the tightly coupled matter-radiation fluid). The characteristics for pressure-free scalar and vector perturbations are timelike curves, moving at zero velocity relative to the matter; while density perturbations with pressure can move at the speed of sound, only tensor perturbations can travel at the speed of light. Thus the true domain that influences us significantly is much less than indicated by the particle horizon. It is the small region round our past world line characterised after decoupling by the comoving scale from which matter coalesced into our galaxy: a present distance of about 1 to 1.95 Mpc,[15] corresponding to an observed angle of about 0.64 arcminutes on the LSS. Before decoupling it would have been limited by the sound horizon [Dodelson, 2003, p. 257] rather than the particle horizon.

2.4.2 Observational limitations

Clearly we cannot obtain any observational data on what is happening beyond the particle horizon; indeed we cannot even see that far because the universe was opaque before decoupling. *Our view of the universe is limited by the* visual horizon, *comprised of the worldlines of furthest matter we can observe — namely, the matter that emitted the CBR at the time of last scattering* [Ellis and Stoeger, 1988]. This occurred at the time of decoupling $t = t_{dec}$ ($z_{dec} \simeq 1100$), and so the visual horizon is characterized by $r = u_{vh}$ where

$$(31) \quad u_{vh} = \int_{t_{dec}}^{t_0} \frac{dt}{S(t)} < u_{ph}.$$

Indeed the LSS delineates our visual horizon in two ways: we are unable to see to *earlier times* than its occurrence (because the early universe was opaque for $t < t_{dec}$), and we are unable to detect matter at *larger distances* than that we see on the LSS (we cannot receive radiation from matter at co-moving coordinate values $r > u_{vh}$). The picture we obtain of the LSS by measuring the CBR from satellites such as COBE and WMAP is just a view of the matter comprising the visual horizon, viewed by us at the time in the far distant past when it decoupled from radiation. The position of the visual horizon is determined by the geometry since decoupling. Visual horizons do indeed exist, unless we live in a small universe, spatially closed with the closure scale so small that we can have seen right around the universe since decoupling. This is a possibility that will be discussed below (Sec. 4.3.1). There is no change in these visual horizons if there was an early inflationary period, for inflation does not affect the expansion or null geodesics during this later period. The major consequence of the existence of visual horizons is that many present-day speculations about the super-horizon structure of the universe — e.g. the chaotic inflationary theory (Sec. 2.6) — are not observationally testable, because one can obtain no definite information whatever about what lies beyond the visual horizon [Ellis, 1975; 1980]. This is one of the major limits to be taken into account in our attempts to test the veracity of cosmological models (Sec. 4.3).

[15]Dodelson [Dodelson, 2003], p. 283; W Stoeger, private communication.

2.5 Theoretical Developments

The cosmological application of Einstein's Theory of Gravitation has also progressed greatly in past decades, as regards exact solutions and generic properties of the field equations; as regards approximate solutions; and in terms of understanding the relationship between them.

2.5.1 Exact solutions and generic properties

Theory initially predicted there must have been a start to the universe, but it was not clear for a long time if this was simply due to the very special exactly isotropic and spatially homogeneous geometry of the standard Friedmann-Lemaître models. It was possible that more realistic models with rotation and acceleration might show the prediction was a mathematical artefact resulting from the idealized models used. The singularity theorems developed by Penrose and Hawking [Hawking and Ellis, 1973; Tipler *et al.*, 1980; Earman, 1999] showed this was not the case: even for realistic geometries, classical gravitational theory predicts a beginning to the universe at a space-time singularity, provided the usual energy conditions were satisfied. This study has led inter alia to a greatly increased understanding of causality and topology of generic universe models [Tipler *et al.*, 1980], including the fact that singularities may have a quite different nature than those in the Robertson-Walker models, for example being anisotropic [Tipler *et al.*, 1980] or of a non-scalar nature [Ellis and King, 1974].

Various classes of exact cosmological solutions are known (Kantowski-Sachs and Bianchi spatially homogeneous but anisotropic models, Tolman-Bondi spherically symmetric inhomogeneous models, and 'Swiss-Cheese' non-analytic models) enabling understanding of dynamical and observational behaviour of more general classes of models than just the FL models [Ellis and van Elst, 1999a]. Dynamical systems studies [Wainwright and Ellis, 1996; Uggla *et al.*, 2003] relate the behaviour of whole classes of anisotropic models in suitable state spaces, enabling identification of generic patterns of behaviour (fixed points, saddle points, attractors, etc.) and hence the relationship between dynamics of higher symmetry and lower symmetry universes. These studies help understanding to what degree the FL models are generic within the families of possible cosmological models, and which models might give observations similar to those in the FL models. In particular they are relevant in considering the possible geometry of the universe at very early or very late times.

2.5.2 Perturbation theory, the gauge issue, and back reaction

Sophisticated perturbation theory has been developed to underlie the theory of structure formation in the expanding universe, examining *the dynamics of perturbed FL models*. The fluid flow in these models can have shear, vorticity, and acceleration, and the Weyl tensor C_{ijkl} (see (1)) is not zero, so that density variations, tidal forces, peculiar velocities, and gravitational waves can be present. Detailed studies use the kinetic theory approximation for matter (electrons, protons, dark matter) and radiation (photons, neutrinos), with their dynamics described by the Boltzmann equation [Dodelson, 2003, Ch. 4]; [Uffink, 2006], interacting with space-time inhomogeneities characterised by a perturbed FL metric. A key issue here is the *gauge problem* — how to choose the background model in the perturbed spacetime [Ellis and Stoeger, 1987]. If this is not properly handled then one may attain apparent perturbation solutions that are pure gauge (they are mathematical rather than physical), so that one can alter the apparent growth rate simply by changing coordinates. The key to handling this is either to keep careful track at all stages of remaining gauge freedom and possible changes of gauge, or (preferably, in my view) to use gauge invariant variables (see [Bardeen, 1980; Ellis and Bruni, 1989; Challinor and Lasenby, 1998]).

Most of the literature on perturbation theory deals with the linear case, but some studies tackle the non-linear regime (e.g. [Langlois and Vernizzi, 2005]), and some consider questions such as the origin of magnetic fields and the causes of galactic rotation. A key problem here is properly relating relativistic analyses of astrophysical dynamics to the Newtonian approaches most often used by astrophysicists (e.g. [Bothun, 1998], pp. 183-222); this is not straightforward.[16] A further unresolved issue is the nature of gravitational entropy [Penrose, 1989b; Ellis, 2002; Penrose, 2004]. Many statements about the nature of entropy in physics textbooks are wrong when gravity is dominant, leading to the spontaneous formation of structures such as stars and galaxies. There is as yet no agreed definition of gravitational entropy that is generally applicable; until there is, cosmological arguments relying on entropy concepts are ill-founded.

The existence of inhomogeneities in the universe raises the issue of fitting and back-reaction. To what degree does the nature of the exactly smooth FL models reflect the geometrical and dynamical nature of more realistic 'lumpy' universe models? [Ellis and Stoeger, 1987]. Inhomogeneities lead to extra terms appearing in the evolution equations for the idealized background models, representing the back-reaction of the perturbations on their dynamics [Ellis, 1984]. These could possibly be dynamically significant [Ellis and Buchert, 2006], but this is a matter of dispute.

2.6 Inflation

Particle physics processes dominated the very early eras, when exotic processes took place such as the condensation of a quark-gluon plasma to produce baryons. Quantum field theory effects were significant then, and this leads to an important possibility: scalar fields producing repulsive gravitational effects could have

[16]Some exact General Relativity results, which must necessarily apply in the Newtonian limit of General Relativity, have no Newtonian analogue; an example is the shear-free theorem applying to pressure-free matter [Ellis, 1967]. The underlying issue is that there are 10 field equations to be satisfied in General Relativity, with 20 integrability conditions (the Bianchi identities), but only one field equation to be satisfied in Newtonian theory (Poisson's equation) together with 4 conservation equations.

dominated the dynamics of the universe at those times. This leads to the theory of the inflationary universe, proposed by Alan Guth [1981; 1997]: if $\mu_{grav} = \mu + 3p/c^2 < 0$, which can happen if a scalar field dominates the dynamics of the early universe, an extremely short period of accelerating expansion will precede the hot big bang era [Blau and Guth, 1987]. This produces a very cold and smooth vacuum-dominated state, and ends in 'reheating': conversion of the scalar field to radiation, initiating the hot big bang epoch. This inflationary process is claimed to explain the puzzles mentioned above (Sec. 2.4.1): why the universe is so special (with spatially homogeneous and isotropic geometry and a very uniform distribution of matter), and also why the space sections are so close to being flat at present (we still do not know the sign of the spatial curvature), which requires very fine tuning of initial conditions at very early times. Inflationary expansion explains these features because particle horizons in inflationary FL models will be much larger than in the standard models with ordinary matter, allowing causal connection of matter on scales larger than the visual horizon, and inflation also will sweep topological defects outside the visible domain.

In more detail: in the case of a single scalar field ϕ with spacelike surfaces of constant density, on choosing u^a orthogonal to these surfaces, the stress tensor has a perfect fluid form with

$$(32)\quad \mu = \frac{1}{2}\dot{\phi}^2 + V(\phi), \;\; p/c^2 = \frac{1}{2}\dot{\phi}^2 - V(\phi),$$

and so

$$(33)\quad \mu + 3p/c^2 = 2\dot{\phi}^2 - 2V(\phi).$$

The slow-rolling case is $\dot{\phi}^2 \ll V(\phi)$, leading to $\mu + p/c^2 = 2\dot{\phi}^2 \simeq 0 \Rightarrow \mu + 3p/c^2 \simeq -2\mu < 0$. This then enables a resolution of the horizon problem in inflationary FL models: if sufficient inflation took place in the early universe, then all the regions from which we receive CBR were causally connected; indeed if the universe began in an inflationary state, or was inflationary with compact spatial sections, there may be no causal horizons at all. The inflationary models also cause initial perturbations to die away, including velocity perturbations, hence explaining the observed smoothness of the universe on large scales. This process is expected to create a universe with spatially very flat sections at late times:

$$(34)\quad \Omega_0 = \Omega_{dm\,0} + \Omega_{\Lambda\,0} \simeq 1 \Leftrightarrow \Omega_k \simeq 0.$$

This theory led to a major bonus: a proposal that initial tiny quantum fluctuations were expanded to such a large scale by inflation that they provided seeds initiating growth by gravitational attraction of large scale structures such as clusters of galaxies. This theory makes clear observational predictions for the spectrum of CBR anisotropies, which have since been spectacularly verified by observations from balloons and satellites, such as WMAP [Spergel *et al.*, 2003]. Thus inflation has provided us with our first coherent theory of structure formation. Inhomogeneities started as quantum fluctuations in the inflationary epoch which are then amplified in physical scale by the inflationary expansion but remain constant in amplitude when larger than the contemporary Hubble scale, leading to Gaussian scale-free perturbations at the start of the HBB era. Starting from these fluctuations, Cold Dark Matter ('CDM') creates potential wells for baryons to fall into, but the radiation (tightly coupled to the electrons and baryons) resists collapse. Gravity wins if the wavelength λ is greater than the *Jean's length* λ_J (which is proportional to the speed of sound [Rees, 1995; Ellis and van Elst, 1999a]). There are acoustic oscillations (sound waves) when $\lambda < \lambda_J$; these oscillations ceased at decoupling, which led to a dramatic decrease in λ_J and the growth of structure by gravitational instability in a 'bottom up' way (Sec. 2.3.6).[17]

A popular version of inflation is *chaotic inflation* [Linde, 1990; Guth, 2001; Susskind, 2005] where inflation ends at different times in different places, so that one ends up with numerous 'pocket universe' (expanding universe domains like the one we see around us, or perhaps very different) all imbedded in a still-inflating universe region and starting at different times, the whole forming a fractal-like structure. It is argued this is an inevitable consequence of the nature of plausible scalar field potentials.

Inflation is not an inevitable conclusion, for there are some alternatives proposed [Hollands and Wald, 2002; Khoury *et al.*, 2001], and the WMAP results can be reproduced by any scenario where Gaussian scale-free perturbations of suitable amplitude occur at the start of the Hot Big Bang era. However inflation is regarded by most cosmologists as the best proposal available for the era prior to the Hot Big Bang epoch, leading to the presence of such perturbations. Nevertheless one should note it is a generic proposal for what happened, rather than a specific physical theory. While a great many possibilities have been proposed (it could for example be an effective field due to higher-order gravity effects, or it could involve multiple scalar fields), at the present time the identity of the proposed inflationary field ('the inflaton') has not been established or linked to any known particle or field. The hoped-for link between early universe dynamics and particle physics is potential rather than real [Earman and Mosterin, 1999]. Detailed studies of the CBR anisotropies and structure formation in conjunction with the observations hope to distinguish between the various possibilities, for example testing whether the spectral index n takes the scale-free value: $n = 1$, or whether rather there is a tilted power spectrum ($n \neq 1$). A unique spectrum of gravitational waves will also be produced at very early times in an inflationary universe, and detection of these waves either directly by proposed gravitational wave detectors or indirectly by measuring the associated curl mode in the CBR polarization will be an important test of inflation, for example determining the ratio r of scalar to tensor perturbations in the early universe [Dodelson, 2003].

[17] This is a highly simplified account; for more detailed versions, see e.g. [Dodelson, 2003; Silk, 2005].

2.7 The very early universe

Quantum gravity processes are presumed to have dominated the very earliest times, preceding inflation. There are many theories of the quantum origin of the universe, but none has attained dominance. The problem is that we do not have a good theory of quantum gravity [Rovelli, 2006], so all these attempts are essentially different proposals for extrapolating known physics into the unknown. A key issue is whether quantum effects can remove the initial singularity and make possible universes without a beginning. Preliminary results suggest this may be so [Bojowald, 2001; Rovelli, 2004; Mulryne *et al.*, 2005].

2.7.1 Is there a quantum gravity epoch?

A preliminary issue is, can there be a non-singular start to the inflationary era, thus avoiding the need to contemplate a preceding quantum gravity epoch? In the inflationary epoch the existence of an effective scalar field leads to a violation of the strong energy condition (8), therefore at first sight it seems that a bounce may be possible preceding the start of the expanding inflationary era and avoiding the inevitability of a quantum gravity epoch.

However a series of theorems suggest that inflationary models cannot bounce: they are stated to be future infinite but not past infinite [Guth, 2001]. This is an important issue, so it is worth looking at it further. There are two major requirements to get a bounce. The Friedmann equation (9) relates the scale factor $S(t)$, curvature constant k, and the effective total energy density $\mu(t)$, which is *defined* by this equation whatever dynamics may be involved (multiple scalar fields, higher order gravity, higher dimensional theories leading to effective 4-dimensional theories, etc.).[18] The Raychaudhuri equation (7) includes the effective total pressure $p(t)$, which again is *defined* by this equation. In this section, a cosmological constant Λ is represented as perfect fluid with $\mu_\Lambda + p_\Lambda/c^2 = 0$. To get a bounce, first one needs the curve $S(t)$ of the scale factor as a function of time to bend up: that is,

$$(35)\quad \frac{\ddot{S}}{S} \geq 0 \Leftrightarrow \mu + 3p/c^2 < 0,$$

which is just a violation of the strong energy condition (8). This is the case if $\mu + p/c^2 = 0$ (a vacuum); and indeed by eqn.(33) it is possible for example for any slow-rolling scalar field. Second, one also needs a time when the scale factor is a minimum. Thus there must be a time t_* such that $\dot{S}(t_*) = 0$. From the Friedmann equation (9),

$$(36)\quad \dot{S}^2(t_*) = 0 \Leftrightarrow \frac{\kappa\mu(t_*)}{3} = \frac{k}{S^2(t_*)}.$$

[18]See [Copeland *et al.*, 2005] for the ways various quantum gravity theories result in modified Friedmann equations.

With $k \leq 0$ this is possible only if $\mu(t_*) < 0$. Even with a scalar field (see eqn.(32)) this can only be achieved by having negative potential energies, which appears to be an unphysical requirement. With $k = +1$ this is possible with $\mu(t_*) > 0$ [Robertson, 1933], which is compatible with ordinary matter.

Thus if you want a bounce in an inflationary universe, it is sensible to look to $k = +1$ inflationary models, which indeed will turn around if a vacuum domain occurs for long enough (curvature will eventually always win over a vacuum as we go back into the past [Ellis *et al.*, 2002b; Ellis *et al.*, 2002a]). The theorems mentioned above do *not* include this case (see [Guth, 2001]); they only consider inflationary universes with $k = 0$ and $k = -1$. And one should note here that although the scale-free $k = 0$ exponential case clearly is the model underlying the way many people approach the problem, it is highly exceptional — it is of zero measure within the space of all inflationary FL models.

Explicit non-singular models can be constructed, the simplest being the de Sitter universe in the $k = +1$ frame (Sec. 2.1.1), which is an exact eternal solution that bounces at a minimum radius S_0. This model has the problem that it does not exit inflation (it corresponds to an exactly constant potential), but variants exist where exit is possible; there are also viable non-singular models that start off asymptotic to the Einstein Static universe in the distant past and avoid the need for a quantum gravity epoch [Ellis and Maartens, 2004]. These models start off in a very special state, precisely because they are asymptotic to the Einstein static universe in the distant past. This is a possible situation. It seems likely that the options for the start of inflation are (i) avoiding the quantum gravity era, but at the cost of having special ('fine tuned') initial conditions, or (ii) having a quantum gravity epoch preceding the inflationary era. Thus a key issue is whether the start of the universe was very special or generic.

2.7.2 Quantum gravity effects: The origin of the universe

Contemporary efforts to explain the beginning of the universe, and the particular initial conditions that have shaped its evolution, usually adopt some approach or other to applying quantum theory to the creation of the universe [Lemaître, 1931]. Many innovative attempts have been made here; as this article focuses on General Relativity and its application to cosmology, and it would be impossible to do justice to the various approaches to quantum cosmology [Rovelli, 2006] without a very much longer article. I will just make a few comments on these approaches.

The attempt to develop a fully adequate quantum gravity approach to cosmology is of course hampered by the lack of a fully adequate theory of quantum gravity, as well as by the problems at the foundation of quantum theory (the measurement problem, collapse of the wave function, etc., see [Isham, 1997; Dickson, 2006; Landsman, 2006]) which can be ignored in many laboratory situations but have to be faced in the cosmological context [Perez *et al.*, 2005]. The various attempts at quantum cosmology each develop in depth some specific aspect of quantum theory that may be expected to emerge from a successful theory of quantum gravity applied to the universe as a whole, being for example based on either (i) the Wheeler-DeWitt equation and the idea of the wave function of the universe, or (ii) on some version of embedding in higher dimensional space time (inspired by string theory), or (iii) an appropriate application of loop quantum gravity. In effect they attempt either

(a) to give a true theory of creation *ex nihilo* [Vilenkin, 1982]; such efforts however cannot truly "solve" the issue of creation, for they rely on some structures or other (e.g. the elaborate framework of quantum field theory and much of the standard model of particle physics) pre-existing the origin of the universe, and hence themselves requiring explanation; or

(b) to describe a self-sustaining or self-referential universe which by-passes the issue of creation, either by

  (b1) originating from an eternally pre-existing state, via the recurring idea of a Phoenix universe [Dicke and Peebles, 1979] (as in Veneziano's 'pre-big bang theory' based on analogues of the dualities of string theory, or self-repeating universes such as the chaotic inflationary models of Linde); creation from fluctuations in some quite different pre-existing structure (e.g. emergence from de Sitter space time; or the 'ekpyrotic universe' initiated by a collision between pre-existing 'branes' in a higher dimensional spacetime); or emerging from an eternal static initial state; or

  (b2) starting from a state with different properties of time than usual (or with an an emergent notion of time): as in the Hartle–Hawking no-boundary proposal [Hawking, 1987; Hawking, 1993], and the Gott causal violation proposal [Gott and Li, 1998] where the universe 'creates itself' and starts normal expansion in the domain without closed timelike lines.

Any of these may be combined with one or other proposals for

(c) an effective ensemble of universes [Tegmark, 2003], realized either

  (c1) in space-time regions that are part of either a larger entangled quantum entity, or are part of a single classical space-time, but are effectively disconnected from each other, or

  (c2) in truly disconnected form.

All of these proposals however are strongly speculative, none being based solidly in well-founded and tested physics, and none being in any serious sense supported by observational evidence. They are all vast extrapolations from the known to the unknown. They may or may not be true. One thing is certain: they can't all be true!

2.8 The concordance model

Observational support for the idea of expansion from a Hot Big Bang epoch is very strong, the linear magnitude-redshift relation for galaxies demonstrating the expansion,[19] with source number counts and the existence of the blackbody CBR being strong evidence that there was indeed evolution from a hot early stage. Agreement between measured light element abundances and the theory of nucleosynthesis in the early universe confirms this interpretation. This basic theory is robust to critical probing. Much present activity attempts to link particle physics interactions during very early stages of the expansion of the universe to the creation of structures by gravitational instability much later, traces of the early seed fluctuations being accessible to us through present day CBR anisotropy patterns. Thus the present dominant cosmological paradigm is *a quantum gravity era of some kind followed by inflation; a hot big bang epoch; decoupling of matter and radiation; and then gravitational instability leading to formation of clusters of galaxies from the seed density perturbations that occur on the LSS.*

Together with supernova data, analysis of the CBR angular anisotropies and in particular their peaks gives a *concordance model* of this kind [Bennet et al., 2003; Tegmark, 2002; Tegmark *et al.*, 2004; Dodelson, 2003; Scott, 2005] that is then confirmed by the statistics of matter clustering [Eisenstein *et al.*, 2005a] together with observations of gravitational lensing and large-scale motions of matter [Silk, 2005]. This model is characterized by specific values for a set of cosmological parameters [Liddle, 2004], in particular

(37) $\Omega_{cdm0} \simeq 0.3$, $\Omega_{\Lambda 0} \simeq 0.7$, $T_{cbr0} = 2.75K$, $H_0 \simeq 65\text{km/sec/mpc}$, $t_0 \simeq 1.4 \times 10^{10}\text{years}$.

Also $\Omega_{bar0} \simeq 0.044$ is the density of baryons, $\Omega_{vis0} \simeq 0.015$ that of luminous matter, and $\Omega_{\nu 0} \simeq 10^{-5}$ that of massless neutrinos, implying $\Omega_0 \simeq 0.3 + 0.7 \simeq 1$ in agreement with the inflationary prediction (34). The sign of k is uncertain, but if the combined evidence of all current observations is taken at face value it is positive, with $\Omega_0 = 1.02 \pm 0.02$ [Spergel *et al.*, 2003]. As noted above, there are some concerns firstly over age issues (see Sec. 2.3.2); secondly concerning the large angle CBR anisotropies (see Sec. 2.3.7); and thirdly regarding details of CDM structure formation at small scales (see Sec. 2.3.6); but none of these issues seems to be crucial at present.

2.8.1 Some misunderstandings

Despite its simplicity, there are some common misconceptions about the standard universe models (cf. [Lineweaver and Davis, 2005]) that can lead to philosophical misunderstandings.

Misconception 1: *The universe is expanding into something.* It is not, as it is all there is. It is just getting bigger, while always remaining all that is. One

[19]The alternative interpretation as gravitational redshifts in a static universe does not work because of the linearity of the observed redshift-distance relation [Ellis *et al.*, 1978].

should note here that a RW universe can be represented as a 4-dimensional curved spacetime expanding in a 5-dimensional flat embedding space time [Robertson, 1933]; however there is no necessity to view the 5-dimensional spacetime in this representation as physically real. Furthermore this embedding is no longer possible when we take perturbations into account; a 10 dimensional flat spacetime is needed for locally embedding a realistic (perturbed) universe model (and to do so globally requires many more dimensions, in general).

Misconception 2: *The universe expands from a specific point, which is the centre of the expansion.* All spatial points are equivalent in these universes, and the universe expands equally about all of them. Every fundamental observer sees exactly the same thing in an exact RW geometry. There is no centre to a FL universe.

Misconception 3: *Matter cannot recede from us faster than light.* It can, at an instant; two distantly separated fundamental observers in a surface $\{t = const\}$ can have a relative velocity greater than c if their spatial separation is large enough [Rothman and Ellis, 1993; Davis and Lineweaver, 2004]. No violation of special relativity is implied, as this is not a local velocity difference, and no information is transferred between distant galaxies moving apart at these speeds. For example, there is presently a sphere around us of matter receding from us at the speed of light;[20] matter beyond this sphere is moving away from us at a speed greater than the speed of light. The matter that emitted the CBR was moving away from us at a speed of about $61c$ when it did so [Rothman and Ellis, 1993].

Misconception 4: *The existence of a preferred RW frame (that in which the universe appears isotropic) contradicts relativity theory, which says all reference frames are equally good.* But this equivalence of frames is true for the equations rather than their solutions. Almost all particular solutions will have preferred world lines and surfaces; this is just a particular example of a *broken symmetry* — the occurrence of solutions of equations with less symmetries than the equations display. This feature is a key theme in modern physics [Brading and Castellani, 2006; Harvey, 2006].

Misconception 5: *The space sections are necessarily infinite if $k = 0$ or -1.* This is only true if they have their 'natural' simply connected topology. If their topology is more complex (e.g. a 3-torus) they can be spatially finite [Ellis, 1971a; Lachièze *et al.*, 1995]. There are many ways this can happen; indeed if $k = -1$ there is an infinite number of possibilities.

Misconception 6: *Inflation implies spatial flatness* $(k = 0 \Leftrightarrow \Omega_k = 1)$ *exactly.* There is nothing in inflationary theory which determines the sign of the spatial curvature. Inflationary universes are very nearly flat at late times; this is very different from being exactly flat (a condition which requires *infinite* fine tuning of initial conditions; if say the two millionth digit in the value of Ω_k is non-zero at

[20]This sphere is not the same as the particle horizon, as is sometimes claimed (see [Rothman and Ellis, 1993]).

any time, then the universe is not spatially flat). Inflationary theory does not have the theoretical teeth required to imply that the universe has exactly flat spatial sections; hence a key issue for cosmology is observationally determining the sign of the spatial curvature, which is potentially dynamically important in both the very early universe [Ellis *et al.*, 2002b; Ellis *et al.*, 2002a] and the late universe (it determines if recollapse is possible, should the dark energy decay away).

2.8.2 Overall

Cosmology has changed from a speculative enterprize into a data-driven science that is part of standard physical theory [Barnett *et al.*, 1996]; a wealth of observations supports this dominant theory [Peebles *et al.*, 1991; Silk, 1997; Perkins, 2005]. Nevertheless some theoretical proposals are being made for the very early stages that have no observational support; and sometimes it may be impossible to ever obtain such support, both as regards the proposed physics and the geometry. Thus in some respects it remains a principle driven enterprise, with observation subordinate to theory.

We now explore the relation between cosmology and philosophy in terms of a series of *Theses* clustered around a set of major *Issues*. One can obtain a synoptic overview of the overall argument by simply considering the full set of *Issues* and *Theses*. They are summarized in the Table at the end.

3 ISSUE A: THE UNIQUENESS OF THE UNIVERSE.

The first and most fundamental issue is that there is only one Universe [Munitz, 1962; McCrea, 1960; Ellis, 1991]. This essential uniqueness of its object of study sets cosmology apart from all other sciences. In particular, the unique initial conditions that led to the particular state of the universe we see were somehow "set" by the time that physical laws as we know them started governing the evolution of both the universe and its contents, whenever that time may be. We cannot alter these unique initial conditions in any way — they are given to us as absolute and unchangeable, even though they are understood as contingent rather than necessary; that is, they could have been different while still being consistent with all known physical laws. The implications are that

Thesis A1: The universe itself cannot be subjected to physical experimentation. *We cannot re-run the universe with the same or altered conditions to see what would happen if they were different, so we cannot carry out scientific experiments on the universe itself.* Furthermore,

Thesis A2: The universe cannot be observationally compared with other universes. *We cannot compare the universe with any similar object, nor can we test our hypotheses about it by observations determining statistical properties of a known class of physically existing universes.*

Where this all becomes of observational relevance is in the idea of *cosmic variance* [Dodelson, 2003, pp. 241, 343]. The theory of structure formation in the early universe makes statistical predictions only (it cannot attempt to predict the specific structures that will actually be formed). Testing the theory compares our universe to a theoretical ensemble of universes, and declares a variance between what is measured in the actual universe and the expected properties based on the ensemble of models. If this variance is small enough, a deviation from expected values is pronounced as a statistical deviation, i.e. of no physical significance — we do not need to explain it any further; if it is large, it needs explanation. This is a key issue for example in the analysis of the CBR anisotropy observations [White *et al.*, 1993; Kamionkowski and Loeb, 1997]. The power spectrum of the CBR as measured by WMAP is less than expected at large angular scales (Sec. 2.3.7). One school of thought claims this is just a statistical fluctuation; another that it needs explanation, and might for example be evidence of a small universe [Luminet *et al.*, 2003; Luminet, 2005]. This debate arises because there is just one universe, and on large angular scales there are just a few measurements that can possibly be made (on small angular scales we can make many measurements and so this uncertainty becomes very small).

Consequent on **A1** and **A2**,

Thesis A3: The concept of 'Laws of Physics' that apply to only one object is questionable. *We cannot scientifically establish 'laws of the universe' that might apply to the class of all such objects, for we cannot test any such proposed law except in terms of being consistent with one object (the observed universe).*

This is insufficient: one observational point cannot establish the nature of a causal relation. Indeed the concept of a 'law' becomes doubtful when there is only one given object to which it applies [Munitz, 1962]. The basic idea of a physical law is that it applies to a set of objects all of which have the same invariant underlying behaviour (as defined by that law), despite the apparent variation in properties in specific instances, this variation resulting from varying initial conditions for the systems on which the law acts. This understanding is tested by physical experiments in which initial conditions for evolution of a set of similar systems are varied, and observations by which the statistical nature of a set of objects of the same broad kind is investigated. Neither is possible in the case of cosmology.

The laws of physics apply locally to the objects in the cosmos, and determine the evolution of the cosmos as a whole when locally applied everywhere with suitable initial/boundary conditions imposed (in the case of the RW models, via the Friedmann equation for example). Apart from this, we cannot establish higher-level effective laws that apply to all universes and determine their structure, as we can at all other levels of the hierarchy of complexity. All that we can do at this level of structure is observe and analyze the one unique object that exists. This is expressed by McCrea as follows: "When we speak of the other solutions of the equations of stellar structure, besides the one we are interested in at the moment, as representing systems that could exist, we mean that they could exist in the universe as we know it. Clearly no such attitude is possible towards the universe itself" [McCrea, 1953].

Since the restriction of a global solution to a local neighborhood is also a solution, we have zillions of "mini-universe" on which to test the laws that control the local nature of the universe. But a mini-universe is not the universe itself; it is a small part of the whole. By examining these "mini-universes" and seeing if they are essentially the same everywhere, we can to some degree check firstly *that the laws of physics are the same everywhere in the universe* (a key feature of all cosmological analysis, cf. Sec. 7.1), and secondly *that the universe is spatially homogeneous* (this is discussed in depth below, see Sec. 4.2.2). But the latter feature is what has to be *explained* by a 'law of the universe'; verifying homogeneity does not explain why it is the case; this comes about because of specific initial conditions, which some suggest are due to hypothesized 'laws of the universe', applicable to the whole rather than to its parts. Finally,

Thesis A4: The concept of probability is problematic in the context of existence of only one object. *Problems arise in applying the idea of probability to cosmology as a whole — it is not clear that this makes much sense in the context of the existence of a single object which cannot be compared with any other existing object.*

But a concept of probability underlies much of modern argumentation in cosmology. Talk of 'fine tuning' for example is based on the use of probability (it is a way of saying something is improbable). This assumes both that things could have been different, and that we can assign probabilities to the set of unrealized possibilities in an invariant way. The issue here is to explain in what sense they could have been different with well-defined probabilities assigned to the different theoretical possibilities, if there is indeed only one universe with one set of initial conditions fixed somehow before physics came into being, or perhaps as physics came into being. We cannot scientifically establish laws of creation of the universe that might determine such initial conditions and resulting probabilities. If we use a Bayesian interpretation, which some suggest can be meaningfully applied to only one object [Garrett and Coles, 1993], the results depend on our 'prior knowledge', which in this case can be varied by changing our initial pre-physics assumptions. Related issues arise concerning the meaning of 'the wave function of the universe', at the heart of quantum cosmology. This wave function gives no unique prediction for any specific single universe.

Two comments on the above. First, *it is useful to distinguish between the experimental sciences — physics, chemistry, microbiology for example — on the one hand, and the historical and geographical sciences — astronomy, geology, evolutionary theory for example, on the other.* It is the former that are usually in mind in discussions of the scientific method. The understanding in these cases is that we observe and experiment on a class of identical or almost identical objects and establish their common behaviour. The problem then resides in just how identical those objects are. Quarks, protons, electrons, are all exactly identical to each other, and so have exactly the same behaviour (indeed this feature underlies well-tested quantum statistics). All DNA molecules, frogs, human beings, and ecosystems are somewhat different from each other, but are similar enough nevertheless that the same broad descriptions and laws apply to them; if this were not so, then we would be wrong in claiming they belonged to the same class of objects in the first place. Water molecules, gases, solids, liquids are in an intermediate category — almost identical, certainly describable reliably by specific physical and chemical laws.

As regards the geographical and historical sciences, here one explicitly studies objects that are unique (the Rio Grande, the continent of Antarctica, the Solar System, the Andromeda galaxy, etc.) or events that have occurred only once (the origin of the Solar System, the evolution of life on Earth, the explosion of SN1987a, etc.). Because of this uniqueness, comment **A1** above applies in these cases also: we can only observe rather than experiment; the initial conditions that led to these unique objects or events cannot be altered or experimented with. However comment **A2** does not apply: at least in principle, there is a class of similar objects out there (other rivers, continents, planetary systems, galaxies, etc.) or similar events (the origin of other galaxies, the evolution of other planetary systems, the explosion of other supernovae, etc.) which we can observe and compare with our specific exemplar, also carrying out statistical analyses on many such cases to determine underlying patterns of regularity; and in this respect these topics differ from cosmology.

If we truly cannot carry out such analyses — that is, if **A2** applies as well in some particular case — then that subject partakes in this respect of the nature of cosmology. One may claim that *the dividing line here is that if we convince ourselves that some large-scale physical phenomenon essentially occurs only once in the entire universe, then it should be regarded as part of cosmology proper*; whereas if we are convinced it occurs in many places or times, even if we cannot observationally access them (e.g. we believe that planets evolved around many stars in other galaxies), then study of that class of objects or events can be distinguished from cosmology proper precisely because there is a class of them to study. The second comment is that some workers have tried to get around this set of problems by essentially *denying the uniqueness of the universe*. This is done by proposing the physical existence of 'many universes' to which concepts of probability can be properly applied (cf. Sec. 2.7.2), envisaged either as widely separated regions of a larger universe with very different properties in each region (as in chaotic inflation for example), as multiple realizations of quantum outcomes, or as an ensemble of completely disconnected universes — there is no physical connection whatever between them — in which all possibilities are realized. We return to this in Sec. 9.2.

4 ISSUE B: THE LARGE SCALE OF THE UNIVERSE IN SPACE AND TIME.

The problems arising from the uniqueness of the universe are compounded by its vast scale in both space and time, which poses major problems for observational cosmology. We therefore need to adduce various Principles in addition to the observations, in order to attain unique models: theory comes in as basis for interpreting observations.

4.1 Observations in a large scale universe

The distance to the nearest galaxy is about 10^6 light years, that is about 10^{24}cm., while the size of the earth is about 10^9cm. The present size of the visible universe is about 10^{10} light years, that is about 10^{28}cm. This huge size relative to our own physical scale (about 10^2cm) places major constraints on our ability to observe distant regions (and certainly prevents us experimenting with them). The uniqueness of cosmology in this respect is that it deals with this scale: the largest with which we can have causal or observational contact.

Thesis B1: Astronomical observations are confined to the past null cone, fading with distance. *We can effectively only observe the universe, considered on a cosmological scale, from one space-time event. Visual observations are possible only on our past light cone, so we are inevitably looking back into the past as we observe to greater distances. Uncertainty grows with distance and time.*

The vast scale of the universe implies we can effectively only view it from one spacetime event ('here and now') [Ellis, 1971a; Ellis, 1975]. If we were to move away from this spatial position at almost the speed of light for say 10,000 years, we would not succeed in leaving our own galaxy, much less in reaching another one; and if we were to start a long-term astronomical experiment that would store data for say 20,000 years and then analyze it, the time at which we observe the universe would be essentially unchanged (because its age is of the order of 10^{10} years: the extra time would make a negligible difference). This is quite unlike other geographic sciences: we can travel everywhere on earth and see what is there. The situation would be quite different if the universe were much smaller. Given its actual scale, such that we are now seeing galaxies whose present distance from us is about 10^9 light years, the effect is as if we were only able to observe the earth from the top of one mountain, and had to deduce its nature from those observations alone [Ellis, 1975].

Because we can only observe by means of particles — photons, massless neutrinos, gravitons — travelling to us at the speed of light, astronomical observations of distant sources and background radiation by telescopes operating at all wavelengths (optical, infrared, ultraviolet, radio, X-ray) are constrained to rays lying in our past light cone. These allow detailed observations (including visual pictures, spectral information, and polarization measurements) of matter as it intersects our past light cone. In observing distant regions, we can also aspire to use neutrino and gravitational wave telescopes, and perhaps cosmic rays, also representing information coming to us at the speed of light or less. However all our detailed data about distant regions is gathered along our past light cone.

As a consequence, three interrelated problems occur in interpreting the astronomical observations. The first is that (because we can only view the universe from one point) *we only obtain a 2-dimensional projection on the sky of the 3-dimensional distribution of matter in the universe.* To reconstruct the real distribution, we need reliable distance measurements to the objects we see. However because of variation in the properties of sources, most are not reliable standard candles or standard size objects to use in calibrating distances, and in these cases we have to study statistical properties of classes of sources to estimate distances.

Second, *we necessarily see distant galaxies and other objects at earlier times in their history* (where their world lines intersect this past light cone).[21] Thus cosmology is both a geographic and a historical science combined into one: we see distant sources at an earlier epoch, when their properties may have been different. As we are looking back in the past, source evolution must be taken into account; their properties at the time they emitted the light may be quite different from their properties now. We can only determine the distances of objects if we understand this evolution; but in practice it is one of the unknowns we have to try to determine (cf. Sec. 4.2.3).

Third, distant sources appear very small and very faint, both because of their physical distance, and because their light is highly redshifted (due to the expansion of the universe). Simply detecting them, let alone determining their characteristics, becomes rapidly more difficult with distance. Furthermore absorption by intervening matter can interfere with light from distant objects. The further back we look, the worse these problems become; thus our reliable knowledge of the universe decreases rapidly with distance [Ellis, 1975].

The situation is however improved by the availability of geological-type data [Hoyle, 1960]; that is, the present-day status of rocks, planets, star clusters, galaxies, and so on, which contains much information on the past history of the matter comprising those objects. Thus we can obtain detailed information on conditions near our past world-line in spacetime [Ellis, 1971a; Ellis, 1975] at very early times if we can interpret this data reliably, for example by relating theories of structure formation to statistical studies of source properties.

Thesis B2: 'Geological' type observations can probe the distant past of our past world line. *Physical and astrophysical observations tell us about conditions near matter world-lines in the far distant past. They can be used also to investigate the far distant past of more distant objects.*

This involves us in physical cosmology: namely the study of the evolution of structures in the universe, tested by comparison with astronomical observation. Particularly useful are measurements of the abundances of elements which resulted

[21]For example we see the Andromeda galaxy as it was two million years ago, long before humans existed on Earth [Silk, 2005].

from nucleosynthesis in the Hot Big Bang, giving us data about conditions long before decoupling (Sec. 2.2.2). If we can obtain adequate quality data of this kind for objects at high redshifts, we can use this to probe conditions very early on in their histories at some distance from our past worldline. Encouraging in this regard is the possibility of determination of element abundances at high redshift [Dodelson, 2003, pp. 11-12]; [Pettini, 1999]).

4.2 Determining Spacetime Geometry: Observational Limits.

The unique core business of observational cosmology is determining the large-scale geometry of everything there is, or at least of everything we can observe.

4.2.1 Direct determination versus theory based approaches

One can go about this in a direct manner: trying to determine the geometry of the universe directly from observations (assuming one has some understanding of the sources observed). The way this can be done (curiously known as the 'inverse approach') has been fully characterized [Kristian and Sachs, 1966; Ellis *et al.*, 1985]; indeed there is an interesting result here, namely

Observational Cosmology Theorem: *The data in principle available on our past null cone from astronomical observations is just necessary and sufficient to determine the space-time geometry on that null cone* [Ellis *et al.*, 1985]. *From this data one can in principle determine the space time in the past of the null cone and, if a no-interference conditions is assumed, to its future.*

However this is difficult to carry out both because of the problem of estimating distances for all observed sources, requiring a knowledge of the nature of the sources (Sec. 4.2.3),[22] and because of the serious difficulty in obtaining some of the needed data (which include apparent distortions of all distant objects, and the transverse velocities of all observed matter). The further we observe down the past light cone, the larger the uncertainty becomes. This direct observational approach, where no prior model is assumed for the space-time geometry, has been pursued to some degree (and in essence underlies for example the observational studies that discovered large-scale structure such as the great walls and voids). Nevertheless it is not widely adopted as an overall approach to cosmology, both because of these observational difficulties, but also because it has little explanatory value; it just tells us what the geometry and matter distribution is, but not why it is of that nature.

The usual option in cosmology proper is rather to use a theory-based approach: we *a priori* assume a model based on a space-time geometry with high symmetry (usually a FL model, see Sec. 2.1), and then determine its essential free parameters from comparison of theoretical relations with astronomical observations (Sec. 2.3.3). Detailed observations of the matter distribution and large-scale velocities

[22]The link between observations and models always requires some theory, and is never direct.

as well as CBR anisotropies then help us determine deviations from the exact model, both statistically (an astrophysical description [Dodelson, 2003]) and in detail (an astronomical description [Ellis and Stoeger, 1987]).

4.2.2 Indirect determination: justifying a Friedmann-Lemaître geometry

The standard models of cosmology are the Friedmann-Lemaître (FL) family of universe models that are exactly spatially homogeneous and isotropic everywhere (Sec. 2.1). They are easy to understand, and have tremendous explanatory power; furthermore their major physical predictions (the existence of blackbody CBR and specific light element production in the early universe) seem confirmed. The issue is, to what degree does observational data uniquely indicate these universe models for the expanding universe geometry? Here one is assuming a large enough averaging scale for spatial homogeneity to be valid; this scale should be explicitly indicated [Ellis, 1984] (it is about 100 Mpc at present [Dodelson, 2003]).[23] These are the background models for cosmology; perturbed FL models then characterize the nature of deviations from the exact FL geometry that are expected on smaller scales (Sec. 2.5.2).

The key feature here is the observed *isotropy about our location* (Sec. 2.3.1). Considered on a large enough angular scale, astronomical observations are very nearly isotropic about us, both as regards source observations and background radiation; indeed the latter is spectacularly isotropic, better than one part in 10^4 after a dipole anisotropy, understood as resulting from our motion relative to the rest frame of the universe, has been removed [Partridge, 1995]. Because this applies to all observations (in particular, there are not major observed matter concentrations in some other universe region), this establishes that in the observable region of the universe, to high accuracy *both the space-time structure and the matter distribution are isotropic about us.* We can easily construct exact spherically symmetric universe models [Bondi, 1947; Ellis and van Elst, 1999a], as indicated by these observations. In general they will be spatially inhomogeneous, with our Galaxy located at or near the centre; this is currently a philosophically unpopular proposal, but is certainly possible. The question is whether we can give convincing observational evidence for spatial homogeneity in addition to the spherical symmetry. Various arguments are used for this purpose.

(a) *The cosmological principle* [Bondi, 1960; Weinberg, 1972]: Just assume spatial homogeneity because it is the simplest case and you don't need anything more complex on the basis of current data. We simply adopt a philosophical principle as the basis of argument. This is essentially an *a priori prescription for initial conditions for the universe* (a universe that initially has a RW geometry will have that geometry at later times, because symmetries of

[23]There exist *hierarchical models* where neither the fluid approximation nor homogeneity is ever attained at any scale because of their fractal nature [de Vaucouleurs, 1970]. The regularity of the observed galactic motions, as evidenced by the (m, z) relations, speaks against these models, as do large-scale observations of the matter distribution [Peebles, 1993a].

the initial data are preserved by the Einstein equations [Hawking and Ellis, 1973]); but it is not usually expressed that way.

(b) *FL observational relations*: If we could show that the source observational relations had the unique FL form (26, 28) as a function of distance, this would establish spatial homogeneity in addition to the isotropy, and hence a RW geometry [Ellis *et al.*, 1985]. This is essentially what is done for example in using number counts towards establishing spatial homogeneity [Hubble, 1936]. However because of Thesis **B1** above, the observational problems mentioned earlier — specifically, unknown source evolution — prevent us from carrying this through: we cannot measure distances reliably enough. Astrophysical cosmology could resolve this in principle, but is unable to do so in practice. Indeed the actual situation is the inverse: *taking radio-source number-count data at its face value, without allowing for source evolution, contradicts a RW geometry.*

In the face of this, the usual procedure is to assume that spatial homogeneity is known in some other way, and deduce the source evolution required to make the observations compatible with this geometric assumption [Ellis, 1975]. It is always possible to find a source evolution that will achieve this [Mustapha *et al.*, 1998]. Thus attempts to observationally prove spatial homogeneity this way fail; indeed an alternative interpretation would be that this data is evidence of spatial inhomogeneity, i.e. that we live in a spherically symmetric inhomogeneous universe where we are situated somewhere near the centre, with the cosmological redshift being partly gravitational, cf. [Ellis *et al.*, 1978] (and conceivably with a contribution to the CBR dipole from this inhomogeneity if we are a bit off-centre). Similarly the supernova data usually understood as implying the existence of a cosmological constant (Sec. 2.3.5) could also be interpreted in this way as evidence of inhomogeneity, without the need for 'dark energy'. Most people regard such proposals as very unappealing — but that does not prove they are incorrect.

(c) *Physical arguments*: One can claim that physical processes such as inflation (Sec. 2.6) make the existence of almost-RW regions highly likely, indeed much more probable than spherically symmetric inhomogeneous regions. This is a viable argument, but we must be clear what is happening here — we are replacing an observational test by a theoretical argument based on a physical process that may or may not have happened (for there is no definitive observational proof that inflation indeed took place). It is strongly bolstered because predictions for the detailed pattern of CBR anisotropy on small scales [Hu and Sugiyama, 1995b], based on the inflationary universe theory, have been confirmed [Perkins, 2005]; but that argument will only become rigorous if it is shown that spherically symmetric inhomogeneous models (with or without inflation) cannot produce similar patterns of anisotropy. But they probably can, because the acoustic oscillations that lead to the characteristic predicted anisotropy patterns in fact take place after inflation, and can equally happen if suitable initial conditions occur without a previous inflationary phase.

What about alternative observational routes? Another proposal is,

(d) *Uniform thermal histories*: the idea is to use the uniformity in the nature of the objects we see in the sky (we see the same types of galaxy at large distances, for example) to deduce they must have all undergone essentially the same thermal history, and then to prove from this homogeneity of thermal histories that the universe must be spatially homogeneous. For example, observations showing that element abundances at high redshift in many directions are the same as locally, are very useful in constraining inhomogeneity by showing that conditions in the very early universe at the time of nucleosynthesis must have been the same at distant locations in these directions [82]. However turning this idea into a proper test of homogeneity has not succeeded so far: indeed it is not clear if this can be done, because some (rather special) counter-examples to this conjecture have been found [Bonnor and Ellis, 1986]. Nevertheless the approach could be used to give evidence against spatial homogeneity: for example, if element abundances were measured to be different at high redshifts in any direction [Pettini, 1999; Sigurdson and Furlanetto, 2005], or if ages of distant objects were incompatible with local age estimates [Jain and Dev, 2005].

Finally the argument for spatial homogenity that is most generally accepted:

(e) *Isotropy everywhere:* If all observers see an isotropic universe, then spatial homogeneity follows [Walker, 1944; Ehlers, 1993; Ellis, 1971a]; indeed homogeneity follows if only three spatially separated observers see isotropy. Now we cannot observe the universe from any other point, so we cannot observationally establish that far distant observers see an isotropic universe. Hence the standard argument is to assume a *Copernican Principle:* that we are not privileged observers. This is plausible in that all observable regions of the universe look alike: we see no major changes in conditions anywhere we look. Combined with the isotropy we see about ourselves, this implies that *all observers see an isotropic universe*, and this establishes a RW geometry [Walker, 1944; Ellis, 1971a; Hawking and Ellis, 1973]. This result holds if we assume isotropy of *all* observations; a powerful enhancement was proved by Ehlers, Geren, and Sachs [Ehlers *et al.*, 1968; Hawking and Ellis, 1973], who showed that if one simply assumes isotropy of freely-propagating radiation about each observer in an expanding universe domain,[24] the result follows from the Einstein and Liouville equations; that is,

[24]This result does not hold in a static universe, for then the radiation temperature depends only on the potential difference between the emitter and observer, hence the radiation is isotropic everywhere even if the universe inhomogeneous, cf. [Ellis *et al.*, 1978].

> **EGS Theorem:** *Exact isotropy of the CBR for every geodesically moving fundamental observer at each point in an expanding universe domain U implies an exact RW geometry in U.*

Thus we may establish spatial homogeneity by assuming a weak Copernican principle: we are not in a privileged position where the CBR just happens to be highly isotropic by chance; hence all comoving observers may be assumed to measure highly isotropic CBR, and the result follows. This is currently the most persuasive observationally-based argument we have for spatial homogeneity.

A problem is that it is an exact result, assuming exact isotropy of the CBR; is the result stable? Indeed it is: *almost-isotropy of freely-propagating CBR for an expanding family of geodesically-moving fundamental observers everywhere in some region proves the universe geometry is almost-RW in that region* [Stoeger *et al.*, 1995]. Thus the result applies to the real universe — provided we make the Copernican assumption that all other observers, like us, see almost isotropic CBR. And that is the best we can do at present. Weak tests of the isotropy of the CBR at other spacetime points come from the Sunyaev-Zel'dovich effect [Goodman, 1995] and from CBR polarization measurements [Kamionkowski and Loeb, 1997], giving broad support to this line of argument but not enough to give good limits on spatial inhomogeneity.

The observational situation is clear:

Thesis B3: Establishing a Robertson-Walker geometry for the universe relies on plausible philosophical assumptions. *The deduction of spatial homogeneity follows not directly from astronomical data, but because we add to the observations a philosophical principle that is plausible but untestable.*

The purpose of the above analysis is not to seriously support the view that the universe is spherically symmetric and inhomogeneous, as is allowed by the observations, but rather to show clearly the nature of the best observationally-based argument by which we can (quite reasonably) justify the assumption of spatial homogeneity.

Accepting this argument, the further question is, *in which spacetime regions does it establish a RW-like geometry*? The CBR we detect probes the state of the universe from the time of decoupling of matter and radiation (at a redshift of about 1100) to the present day, within the visual horizon. The argument from CBR isotropy can legitimately be applied for that epoch. However, it does not necessarily imply isotropy of the universe at much earlier or much later times, because there are spatially homogeneous anisotropic perturbation modes that are unstable in both directions of time; and they will occur in a generic situation. Indeed, if one examines the Bianchi (spatially homogeneous but anisotropic) universes, using the powerful tools of dynamical systems theory, one can show that *intermediate isotropisation* can occur [Wainwright and Ellis, 1996; Wainright *et al.*, 1998]: *despite being highly anisotropic at very early and very late times, such models can mimic a RW geometry arbitrarily closely for an arbitrarily long time*, and hence can reproduce within the errors any set of FL-like observations. We can obtain strong limits on the present-day strengths of these anisotropic modes from CBR anisotropy measurements and from data on element abundances, the latter being a powerful probe because (being of the 'geological' kind) they can test conditions at the time of element formation, long before decoupling. But however low these observational limits, anisotropic modes can dominate at even earlier times as well as at late times (long after the present). If inflation took place, this conclusion is reinforced: it washes out any information about very early universe anisotropies and inhomogeneities in a very efficient way.

As well as this time limitation on when we can regard homogeneity as established, there are major spatial limitations. The above argument does not apply far outside the visual horizon, for we have no reason to believe the CBR is highly isotropic there. Indeed if chaotic inflation is correct, conditions there are not the same.

4.2.3 Determining the RW parameters

Given that a RW geometry is a good description of the observable universe on a large scale, the further issue is what are the best-fit parameters that characterize it, selecting the specific universe we observe from the family of all FL models (Sec. 2.1). Important observational issues are:

- Determining the Hubble parameter H_0, which sets the overall scale of the observed universe region.

- Determining the trio of the density parameter Ω_0, deceleration parameter q_0, and cosmological constant Λ (or equivalently the density parameter Ω_Λ), which are the major defining characteristics of a specific FL model. The CBR data, supernova observations, deep number counts, source covariance functions, velocity measurements, and gravitational lensing observations can determine these quantities.

- Determining the sign of the curvature k, showing whether the universe has closed spatial sections and also whether it is possible for it to recollapse in the future or not. Analyses of the observations should always attempt to determine this sign, and not assume that $k = 0$ (as is often done) [Wright, 2006].

- Various parameters are used to characterize the nature of dark matter (Sec. 2.3.6) and dark energy (Sec. 2.3.5). As their dynamics is unknown, these too have to be determined observationally.

We only obtain good estimates of these quantities by the observational relationships characterized above (Sec. 2.3.3) using statistical analysis of the classes of objects we observe. Problems arise because of our lack of adequate theories of their historical development.

Thesis B4: Interpreting cosmological observations depends on astrophysical understanding. *Observational analysis depends on assessing a variety of auxiliary functions characterizing the sources observed and the observations made. These introduce further parameters that have to be observationally or theoretically determined, allowing considerable freedom in fitting specific models to the observations. Physical cosmology aims to characterize perturbed FL models (which account for structure formation) rather than just the background exactly smooth FL models; this introduces further parameters to be determined.*

It is useful here to distinguish between methods aimed at determining the properties of the background (zeroth order) FL model directly, and those aimed at determining properties of the perturbations of these models [Tegmark, 2002]. Methods for determining the parameters of the background model (Sec. 2.1) depend on assuming properties of the distance indicators used (galaxies, radio sources, etc.). They will have their own properties (brightness profiles, luminosities, physical sizes, spectra, etc.) and dynamical evolution; but these are often not well understood, and will have to be represented in a parametric way (e.g. by parameters describing luminosity evolution). In each case we end up assuming important aspects of the astrophysics and evolutionary histories of the objects observed, which are not part of the cosmological model proper. The statistical properties of the sources observed are also characterized by parametrized functions (e.g. the luminosity function characterizing the numbers of galaxies in each luminosity class) that have to be known in order to analyze the observations. This situation is an exemple of Lakatos' view of how scientific programmes work, with a belt of auxiliary hypotheses interposing between the core theoretical proposal and the data used to test it [Lakatos, 1980]. This makes the analysis rather model-dependent, where the models are only indirectly related to the background model — their explanation is the aim of astrophysics rather than cosmology. Thus if observational results disagree with a particular cosmological model, one can always claim it is the understanding of the auxiliary hypotheses that is at fault rather than the model being proposed [Lakatos, 1980].

By contrast, many of the methods of estimating Ω_0 (and to some degree Λ) depend on studying the growth and nature of inhomogeneities in the universe, that is they investigate perturbed FL models (Sec. 2.5.2), whose properties of course depend on the background model, but introduce a whole set of further functions and parameters describing the perturbations [Dodelson, 2003], for example the angular correlation function for matter (or its Fourier transform, the 2-dimensional power spectrum), the power spectrum of density fluctuations [Tegmark, 2002], red-shift space correlation functions [Peebles, 1993a; Eisenstein *et al.*, 2005a], and correlation function for velocities [Dodelson, 2003]. Associated parameters include a scalar *spectral index* (characterizing the spectrum of physical sizes of inhomogeneities), the *bias parameter* b (expressing how galaxy formation is biassed towards density peaks in the inhomogeneities [Dodelson, 2003, p. 280]) and the *initial fluctuation magnitudes* Q (the seeds for structure formation). Determining these parameters is part of the task of cosmology proper: to fully characterize the perturbed cosmological model, *we aim to determine both the background parameters and the quantities describing the perturbations.* Model selection then depends on the parameters used to describe them — what is assumed known, and what is to be determined [Liddle, 2004; Scott, 2005]. For example, standard inflationary theory predicts a scale-invariant spectrum of Gaussian perturbations; do we test that assumption, or take it for granted? This comes up in the issue of what 'priors' are assumed when conducting statistical tests.

4.2.4 Consistency tests

A key question for cosmology is *what kinds of observations provide critical tests of the standard FL models.* If there were no observations that could disprove them, the subject would be of questionable scientific status. An important such test is obtaining *estimates of the age of the universe* t_0, which is dependent on H_0, Ω_0, and Λ, and comparing them with estimates of the ages of objects in the universe (determined on astrophysical grounds):

Thesis B5: A crucial observational test for cosmology is that the age of the universe must be greater than the ages of stars. *The tension between the age of the universe and ages of stars is one area where the standard models are vulnerable to being shown to be inconsistent, hence the vital need to establish reliable distance scales, basic to estimates of both* H_0 *and the ages of stars, and good limits on* Λ. *Other consistency tests help confirm the standard model and consolidate cosmology's standing as an empirical science.*

At present this age issue is acceptable for local objects, because of a recent revision of our distance scale estimates [Harris *et al.*, 1998], assisted by data that Λ is positive [Perlmutter *et al.*, 1998]; but continued vigilance is needed on this front, particularly as there are indications of problems for high redshift objects [Jain and Dev, 2005]. If this ever became serious we might have to resort to spherically symmetric inhomogeneous models rather than spatially homogeneous models, with the 'bang time' (characterizing the start of the universe) dependent on distance from us [Mustapha *et al.*, 1998].

Note that this issue is crucially unlike the case of the large angle CBR anisotropies (Sec. 2.3.7): *the low CBR anisotropies at large angular scales can as a last resort be dismissed as a statistical fluke; the age issue cannot.* It is to do with the internal consistency of individual cosmological models, not with probabilities. Thus it is a plus for cosmology that the age issue exists. Other consistency tests include

- Showing that the CBR temperature T_{cbr} varies with redshift according to $T_{cbr} = 2.75\,(1+z)$ [Meyer, 1994];

- Confirming that helium abundances are consistent with a primordial value of 25% at large distances (high redshifts) in all directions [Dodelson, 2003, pp. 11-12]; also [Pettini, 1999; Sigurdson and Furlanetto, 2005]; and

- Checking that there is a 2% number count dipole parallel to the CBR dipole for all cosmological sources [Ellis and Baldwin, 1984].

4.3 The hidden universe

If we do not live in a small universe (Sec. 4.3.1), the further essential point is that the region of the universe we can observe is restricted, firstly because we cannot see to earlier times than the LSS (the universe was opaque before then (see Sec. 2.2)), and secondly because a finite time has elapsed since the universe became transparent to radiation, and light can only have travelled a finite distance in that time. As no signal can travel to us faster than light, we cannot receive any information from galaxies more distant than our visual horizon [Ellis and Stoeger, 1988]. The most distant matter we can observe is that which emitted the CBR (Sec. 2.4.2).

Thesis B6: Observational horizons limit our ability to observationally determine the very large scale geometry of the universe. *We can only see back to the time of decoupling of matter and radiation, and so have no direct information about earlier times; and unless we live in a 'small universe', most of the matter in the universe is hidden behind the visual horizon. Conjectures as to its geometry on larger scales cannot be observationally tested. The situation is completely different in the small universe case: then we can see everything there is in the universe, including our own galaxy at earlier times.*

The key point here is that unless we live in a small universe, *the universe itself is much bigger than the observable universe.* There are many galaxies — perhaps an infinite number — at a greater distance than the horizon, that we cannot observe by any electromagnetic radiation. Furthermore no causal influence can reach us from matter more distant than our particle horizon — the distance light can have travelled since the creation of the universe, so this is the furthest matter with which we can have had any causal connection [Rindler, 1956; Hawking and Ellis, 1973; Tipler *et al.*, 1980]. We can hope to obtain information on matter lying between the visual horizon and the particle horizon by neutrino or gravitational radiation observatories; but we can obtain no reliable information whatever about what lies beyond the particle horizon. We can in principle feel the gravitational effect of matter beyond the horizon because of the force it exerts (for example, matter beyond the horizon may influence velocities of matter within the horizon, even though we cannot see it). This is possible because of the constraint equations of general relativity theory, which are in effect instantaneous equations valid on spacelike surfaces.[25] However we cannot uniquely decode that signal to determine what matter distribution outside the horizon caused it: a particular velocity field might be caused by a relatively small mass near the horizon, or a much larger

[25]They are valid at any late time in a solution of the EFE because they were valid initially — the initial data must satisfy constraint equations — and once they are satisfied, the constraints are preserved by the dynamic field equations.

mass much further away [Ellis and Sciama, 1972]. Claims about what conditions are like on very large scales — that is, much bigger than the Hubble scale — are unverifiable [Ellis, 1975], for we have no observational evidence as to what conditions are like far beyond the visual horizon. The situation is like that of an ant surveying the world from the top of a sand dune in the Sahara desert. Her world model will be a world composed only of sand dunes — despite the existence of cities, oceans, forests, tundra, mountains, and so on beyond her horizon.

It is commonly stated that if we live in a low-density universe and the cosmological constant vanishes, the universe has infinite spatial sections. However this deduction only applies if firstly the RW-like nature of the universe within the past light cone continues to be true indefinitely far outside it, and secondly the space sections have their 'natural' simply-connected topology — and there is no way we can obtain observational evidence that these conditions are both true. In contrast to this, in chaotic inflationary models (Sec. 2.6), it is a definite prediction that the universe will not be like a RW geometry on a very large scale — rather it will consist of many RW-like domains, each with different parameter values, separated from each other by highly inhomogeneous regions outside our visual horizon [Linde, 1990], the whole forming a fractal-like structure. This prediction is just as untestable as the previously prevalent assumption (based on a Cosmological Principle) that the universe is RW-like on such scales [Bondi, 1960; Weinberg, 1972]. Neither can be observationally confirmed or denied. The same issue arises in an even more extreme form in relation to the idea of a multiverse. We return to this below, see Sec. 9.2.

4.3.1 Small universes

There is one case where this kind of spatial observational limit does not obtain. This is when a *Small Universe* occurs, that is, a universe which closes up on itself spatially for topological reasons [Ellis, 1971b], and does so on such a small scale that we have seen right round the universe since the time of decoupling. Then we can see all the matter that exists, with multiple images of many objects occurring [Ellis and Schreiber, 1986]. This possibility is observationally testable by examining source statistics, by observation of low power in the large angle CBR anisotropies, and by detecting identical temperature variation on various circles in the CBR sky [Lachièze *et al.*, 1995]. There are weak hints in the observed CBR anisotropies (the lack of power on large angular scales) that this could actually be the case [Luminet *et al.*, 2003; Luminet, 2005], but this is not solidly confirmed. Checking if the universe is a small universe or not is an important task; the nature of our observational relationship to the universe is fundamentally different if it is true [Ellis and Schreiber, 1986].

4.4 The observed universe

The observable part of the universe (i.e. back to the visual horizon) is strictly limited, and we have already seen most of it. We can only observe distant objects by electromagnetic radiation at all wavelengths, by neutrinos, and by gravitational waves. We already have very complete broad coverage of the entire sky by electromagnetic observations at all wavelengths right back to the surface of last scattering, which is the limit of what will ever be observable by electromagnetic radiation. Detailed observations (such as the Hubble Deep Field) are available for restricted domains in angle and depth. Detailed observations at suitable wavelengths are beginning to discern what lies behind the Milky Way, which tends to obscure a substantial fraction of the sky. It is unlikely there are many new astronomical phenomena undiscovered in this observable region, although it will be crucial determining more detailed features of the phenomena we have already discovered (e.g. the nature of dark matter and dark energy).

Thesis B7: We have made great progress towards observational completeness. *We have already seen most of the part of the universe that is observable by electromagnetic radiation. It is plausible that not many new astronomical phenomena remain to be discovered by us observationally; we will determine more details (so understanding more about what we have seen) and see more objects, but not discover many new kinds of things.*

Indeed Harwit [1984] has used the multiplicity of discovery of specific astronomical phenomena to estimate how many new essentially different such phenomena there are still waiting to be discovered.

Neutrinos and gravitational waves will in principle allow us to peer back to much earlier times (the time of neutrino decoupling and the quantum gravity era respectively), but are much harder to observe at all, let alone in useful directional detail. Nevertheless the latter has the potential to open up to us access to eras quite unobservable in any other way. Maybe they will give us unexpected information on processes in the very early universe which would count as new features of physical cosmology.

5 ISSUE C: THE UNBOUND ENERGIES IN THE EARLY UNIVERSE

The analogous problems for physical cosmology arise because energies occurring in the Hot Big Bang early universe phase (Sec. 2.2) are essentially unbounded, so the highest energies we can attain in particle accelerators cannot reach the levels relevant to very early times. The uniqueness of cosmology in this regard is that it is the only science contemplating spacetime regions that have experienced such high energies, and with which we are in intimate causal contact despite the huge timescales involved — indeed events at those early times determined much of what we see around us today. The nuclear reactions underlying nucleosynthesis are well understood, and their cross-sections reasonably well-known; the processes of baryosynthesis and quark-gluon recombination are reasonably understood and are on the border of being testable; but physical processes relevant at earlier times are inaccessible to testing by laboratory or accelerator-based experiment. The *Physics Horizon* by definition separates those aspects of physics we can hope to test by high-energy experiments on Earth or in the Solar System, from those where it is reasonable to expect no such test will ever be possible:

Thesis C1: The Physics Horizon limits our knowledge of physics relevant to the very early universe. *We cannot experimentally test much of the physics that is important in the very early universe because we cannot attain the required energies in accelerators on Earth. We have to extrapolate from known physics to the unknown and then test the implications; to do this, we assume some specific features of known lower energy physics are the true key to how things are at higher energies. We cannot experimentally test if we have got it right.*

Note that this is independent of the issue of setting of initial conditions for the universe, considered below, see Sec. 6.2: the problem arises after the initial conditions have been set and the universe is running according to invariable physical laws. We cannot be confident of the validity of the physics we presuppose then. Rather than using known physics to predict the evolution of the universe, *we end up testing proposals for this physics by exploring their implications in the early universe*, which is the only 'laboratory' where we can test some of our ideas regarding fundamental physics at the highest energies [Yoshimura, 1988]; this is particularly true in the case of quantum gravity proposals. The problem is we cannot simultaneously do this and also carry out the aim of physical cosmology, namely predicting the evolution of the early universe from known physical theory.

Our understanding of physics at those times has of necessity to be based on extrapolation of known physics way beyond the circumstances in which it can be tested. The trick is to identify which features are the key to use in that extrapolation: for example, variational principles, broken symmetries and phase changes, duality invariance, entropy limits are candidates. If we confirm our guesses for the relevant physics by their satisfactory implications for the early universe, tested in some suitable way, then this is impressive progress; but if this is the *only* way we can test the proposed physics, the situation is problematic. If the hypothesis solves only the specific issues it was designed to solve in the early universe and nothing else, then in fact it has little explanatory power, rather it is just an alternative (perhaps theoretically preferable) description of the known situation. One obtains positive observational support for a particular proposal for the relevant physics only if it predicts multiple confirmed outcomes (rather than just one), for example predicting particles that are then confirmed to exist in a laboratory, so that a single hypothesis simultaneously solves several different observational issues. Some of the options may be preferred to others on various theoretical grounds; but one must distinguish this from their having observational support. They lack physical power if they have no other testable consequences. A particular example is the inflationary universe proposal (Sec. 2.6): the supposed inflaton field underlying an inflationary era of rapid expansion in the early universe [Guth, 1981; Gibbons *et al.*, 1983; Kolb and Turner, 1990; Guth, 1997] has not been identified, much less shown to exist by any laboratory experiment. Because this field ϕ is unknown, one can assign it an arbitrary potential $V(\phi)$, this arbitrariness reflecting our inability to experimentally determine the relevant behaviour. It can be shown that virtually any desired scale evolution $S(t)$ of the universe can be attained by suitable choice of this potential [Ellis and Madsen, 1991]; and also almost any desired perturbation spectrum can be obtained by a (possibly different) suitable choice [Lidsey *et al.*, 1997]. Indeed in each case one can run the mathematics backwards to determine the required potential $V(\phi)$ from the desired outcome (Sec. 9.3.1 below). The mathematical existence of such a theoretical potential of the desired form for cosmological purposes does not by itself prove a particle or field exists with that effective potential.

Thesis C2: The unknown nature of the inflaton means that inflationary universe proposals are incomplete. *The promise of inflationary theory in terms of relating cosmology to particle physics has not been realized. This will only be the case when the nature of the inflaton has been pinned down to a specific field that experiment confirms or particle physics requires to exist.*

The very impressive achievement of inflation is that the predicted CBR anisotropy spectrum is verified and agrees with the matter power spectrum [Eisenstein *et al.*, 2005a]; but that prediction depends only on the physics from the era of tight coupling of matter and radiation to the present day, given a suitable initial fluctuation spectrum in the early universe, rather than on the specific hypothesis of an inflationary origin for that spectrum. The true clincher would be if properties of an inflationary field were predicted from the cosmology side and then confirmed in the laboratory; indeed that would count as one of the great feats of theoretical physics. This may not happen however because of the experimental problems focused on here, arising because we cannot reproduce on Earth all the conditions relevant to very early cosmology.

One key application where this issue becomes significant is in respect of the chaotic inflation theory (Sec. 2.6). As remarked above, see Sec. 4.3, its geometric predictions are observationally unverifiable. It would nevertheless be a good physical prediction if it was a more or less inevitable outcome of known and tested underlying physics. However this is not the case: the proposed underlying physics is not experimentally tested, indeed it is not even uniquely defined or associated with any specific known physical particle or field. The claim that it inevitably follows from string theory [Susskind, 2005] suffers from the problem that string theory is not a well-defined or tested part of physics.

6 ISSUE D: EXPLAINING THE UNIVERSE — THE QUESTION OF ORIGINS.

This is the unique core business of physical cosmology: explaining both why the universe has come into existence and evolved to the present very high-symmetry FL geometry on large scales, and how structures come into existence on smaller scales.

6.1 Start to the universe

Did a start to the universe happen? If so, what was its nature? This has been discussed above (Sec. 2.7.2), and the issue is unresolved. The major related question is whether the process of expansion only happens once in the life of the Universe, or occurs repeatedly. The first option is the standard model, where the entire evolution of the Universe is a once-off affair, with all the objects we see, and indeed the Universe itself, being transient objects that will burn out like dead fireworks after a firework display. In this case everything that ever happens occurs during one expansion phase of the Universe (possibly followed by one collapse phase, which could occur if $k = +1$ and the present 'dark energy' field dies away in the future). This evolution might have a singular start at a space-time singularity; a beginning where the nature of time changes character; a non-singular bounce from a single previous collapse phase; or a start from a non-singular static initial state [Mulryne *et al.*, 2005]. An alternative is that many such phases have occurred in the past, and many more will occur in the future; the Universe is a *Phoenix Universe* [Dicke and Peebles, 1979], new expansion phases repeatedly arising from the ashes of the old. While the idea of one or more bounces is an old one [Tolman, 1934], actual mechanisms that might allow this bounce behaviour have not yet been elucidated in a fully satisfactory way. A variant is the chaotic inflation idea of new expanding universe regions arising from vacuum fluctuations in old expanding regions, leading to a universe that has a fractal-like structure at the largest scales, with many expanding regions with different properties emerging out of each other in a universe that lasts forever (Sec. 2.6).

As discussed above, see Sec. 2.7.1, it is possible (if the universe has positive spatial curvature) that the quantum gravity domain can be avoided and there was no start to the universe; however this probably requires special initial conditions [Ellis and Maartens, 2004]. If a quantum gravity domain indeed occurred, we cannot come to a definite conclusion about whether there was a creation event or not because we do not know the nature of quantum gravity, nor how to reliably apply it in the cosmological context where the issue of initial conditions arises. Loop quantum gravity suggests the universe may be singularity-free [Bojowald, 2001], with bounces or a non-singular start, but that theory is unconfirmed. Tested physics cannot give a decisive answer; it is possible that *testable* physics also cannot do so.

Thesis D1: An initial singularity may or may not have occurred. *A start to the universe may have occurred a finite time ago, but a variety of alternatives are conceivable: eternal universes, or universes where time as we know it came into existence in one or another way. We do not know which actually happened, although quantum gravity ideas suggest a singularity might be avoided.*

This is a key issue in terms of the nature of the universe: a space-time singularity is a dramatic affair, where the universe (space, time, matter) has a beginning and all of physics breaks down and so the ability to understand what happens on a scientific basis comes to an end. However eternal existence is also problematic, leading for instance to the idea of Poincaré's eternal return: everything that ever happened will recur an infinite number of times in the future and has already occurred an infinite number of times in the past [Barrow and Tipler, 1984]. This is typical of the problems associated with the idea of infinity (discussed further below, see Sec. 9.3.2). *It is not clear in the end which is philosophically preferable: a singularity or eternal existence.* That decision will depend on what criteria of desirability one uses (such criteria are discussed below, see Sec. 8.1).

6.2 The issue of initial conditions

While occurrence of an initial singularity is striking in that it is a start to physics and spacetime as well as matter, whether it occurred or not is in a sense irrelevant to the key issue of what determined the nature of the universe:

Thesis D2: Testable physics cannot explain the initial state and hence specific nature of the universe. *A choice between different contingent possibilities has somehow occurred; the fundamental issue is what underlies this choice. Why does the universe have one specific form rather than another, when other forms consistent with physical laws seem perfectly possible? The reasons underlying the choice between different contingent possibilities for the universe (why one occurred rather than another) cannot be explored scientifically. It is an issue to be examined through philosophy or metaphysics.*

Even if a literal creation does not take place, as is the case in various of the present proposals, this does not resolve the underlying issue of what determined why the universe is the way it is, given that it could presumably have been otherwise. If the proposal is evolution from a previous eternal state — Minkowski space for example — then why did that come into existence, and why did the universe expansion as a bubble from that vacuum start when it did, rather than at some previous time in the pre-existent eternity? Whenever it started, it could have started before! Some attempts involve avoiding a true beginning by going back to some form of eternal or cyclic initial state, for example Tolman's series of expansion and collapse cycles [Tolman, 1934], proposals for creation of the universe as a bubble formed in a flat space-time [Tryon, 1973], Linde's eternal chaotic inflation [Linde, 1990], Veneziano's re-expansion from a previous collapse phase [Ghosh *et al.*, 1998], the ekpyrotic universe proposal [Khoury *et al.*, 2001], and theories involving foundational limits on information through a "holographic principle" [Susskind and Lindesay, 2004]. These do not avoid the ultimate problem; it can be claimed they simply postpone facing it, for one now has to ask all the same questions of origins and uniqueness about the supposed prior state to the Hot Big Bang expansion phase. The Hartle-Hawking 'no-boundary' proposal [Hawking, 1993] avoids the initial singularity because of a change of space-time signature, and so gets round the issue of a time of creation in an ingenious way; and Gott's causality violation in the early universe [Gott and Li, 1998] does the same kind of thing in a different way. Such proposals cannot overcome the ultimate existential question: *Why has* *one specific state occurred rather than any of the other possibilities? How was it decided that this particular kind of universe would be the one actually instantiated?* This question cannot be solved by physics alone, unless one can show that only one form of physics is self-consistent; but the variety of proposals made is evidence against that suggestion.

The explanation of initial conditions has been the aim of the family of theories one can label collectively as 'quantum cosmology' [Hawking, 1993; Gott and Li, 1998; Gibbons *et al.*, 2003]; however as discussed earlier, here we inevitably reach the limits to what the scientific study of the cosmos can ever say — if we assume that such studies must of necessity involve an ability to observationally or experimentally check our theories. No physical experiment at all can help here because of the uniqueness of the universe, and the feature that no spacetime exists prior to (in a causal sense) such a beginning; so brave attempts to define a 'physics of creation' stretch the meaning of 'physics'. Prior to the start (if there was a start) physics as we know it is not applicable and our ordinary language fails us because time did not exist, so our natural tendency to contemplate what existed or happened 'before the beginning' is highly misleading — there was no 'before' then, indeed there was no 'then' then! Talking as if there was is commonplace, but quite misleading in trying to understand a scientific concept of 'creation' [Grunbaum, 1989]. We run full tilt into the impossibility of testing the causal mechanisms involved, when physics did not exist. No experimental test can determine the nature of any mechanisms that may be in operation in circumstances where even the concepts of cause and effect are suspect. This comes particularly to the fore in proposing 'laws of initial conditions for the universe' — for here we are apparently proposing a theory with only one object. Physics laws are by their nature supposed to cover more than one event, and are untestable if they do not do so (Sec. 3).

6.3 Special or general

The present state of the universe is very special. Explanation of the present large-scale isotropy and homogeneity of the universe means determining the dynamical evolutionary trajectories relating initial to final conditions, and then essentially either *explaining initial conditions*, where we run into difficulties (Sec. 6.2), or *showing they are irrelevant*. The issue raised is whether the universe started off in a very special geometrical state:

Thesis D3: The initial state of the universe may have been special or general. *Whether there was generality or speciality of geometrical initial conditions for the universe is a key question. It seems likely that the initial state of the observed part of the universe was not generic.*

The assumption that the universe is geometrically special was encoded in the Cosmological Principle, taken as a founding principle in cosmology until the 1960's, i.e. as an 'explanation' of special initial conditions [Bondi, 1960; Weinberg, 1972]. Then Misner introduced the chaotic cosmology programme [Misner, 1968], based on the idea of a universe with generic initial conditions being isotropised at later times by physical processes such as viscosity, making initial conditions irrelevant. This concept of isotropisation then became central to the inflationary family of theories (Sec. 2.6), with the underlying assumption being that 'fine tuning' of initial conditions is unphysical and to be avoided. Both programmes are however only partially successful: one can explain a considerable degree of isotropisation and homogenization of the physical universe by either process, but this will not work in all circumstances. Inflation can get rid of much anisotropy [Wald, 1983] but inhomogeneity must be restricted if inflation is to succeed in producing a universe like that we see today, and the success of inflation in solving the horizon issue for FL models — where exact homogeneity exists to start with — will not necessarily be replicated in anisotropic models. Universes that are initially too anisotropic may never inflate, and the horizon problem may not be solved in such models if they do;[26] and only rather special states lead to ordinary thermodynamics [Penrose, 1989a; Penrose, 2004; Wald, 2005; Carroll and Chen, 2005], which is taken to be true in inflationary physics.

Inflation can only be guaranteed to succeed if initial conditions are somewhat restricted; some degree of geometric speciality must have occurred at the start of the observed region of the universe. This special domain might possibly occur within the context of a much larger universe domain where conditions vary randomly, and only isolated regions lead to inflation and eventually domains such as that we see around us; attractive as this may be, it is an untestable hypothesis (essentially a version of the multiverse proposal, see Sec. 9.2).

Special initial conditions (which inflation proposes to explain) might have just occurred that way. The ultimate issue is that *we have no proof as to whether initial conditions for the universe were special or general; either could have occurred.* If we state these conditions must have been general, we are making a philosophical claim, for it is not a provable physical statement. Part of the problem is that we have no agreed measure on the space of possible universes; what seems special or general depends on the choice of such a measure.

7 ISSUE E: THE UNIVERSE AS THE BACKGROUND FOR EXISTENCE

The universe provides the environment for all of science, by determining the initial conditions within which all physical laws are constrained to operate, thus setting boundary conditions for all local physics. Together with suitable equations of state for the matter or structural equations for complex systems, these determine the nature of physical outcomes. The uniqueness of cosmology lies in that it considers the origin of such conditions.

[26]Most inflationary studies show only that the *geometric* horizon problem is solved in the very special RW geometries; but there is no *physical* horizon problem in those geometries, for they are by by assumption spatially homogeneous and isotropic *ab initio*.

7.1 Laws and boundary conditions

A fundamental assumption underlying physical cosmology is the idea that *the laws of physics are the same everywhere in the physical universe*: those we determine in a laboratory here and now will be the same as apply at very distant places (e.g. determining the astrophysics of qso's at redshift $z = 6$), at very early times (e.g. at the time of nucleosynthesis), and at very late times. Without this assumption, explanatory theories have no solid foundation. However because of the uniqueness of the universe discussed above (see Sec. 3), unlike the rest of physics where the distinction is clear and fundamental, in the cosmological context the distinction between laws and boundary conditions becomes blurred.

Thesis E1: Physical laws may depend on the nature of the universe. *We have an essential difficulty in distinguishing between laws of physics and boundary conditions in the cosmological context of the origin of the universe. Effective physical laws may depend on the boundary conditions of the universe, and may even vary in different spatial and/or temporal locations in the cosmos.*

Because we cannot vary the initial conditions in any way, as far as we are concerned they are necessary rather than contingent — so the essential distinction between initial conditions and laws is missing. The distinction is clear once the cosmos has come into existence — but we are concerned with 'prior' conditions associated with the creation of the cosmos and the very existence of physical laws. Certainly any proposal for distinguishing between laws of nature and boundary conditions governing solutions to those laws is untestable in this context. Given the feature that the universe is the unique background for all physics, it is therefore not far-fetched to suggest that it is possible the cosmos influences the *nature* of local physical laws, rather than just their initial conditions [Ellis and Sciama, 1972; Ellis, 2002]. This has been examined over many decades in three specific cases.

(a) *Varying 'constants'*: It might be that there is a time variation in physical constants of nature [Barrow, 2003] related to the expansion of the universe, as proposed in the case of the gravitational constant G by Dirac [Dirac, 1938], developed in depth by Jordan and then Brans and Dicke [Brans and Dicke, 1961]. Such proposals must be consistently developed in relation to the rest of physics and should be related to dimensionless constants, as otherwise they may simply be disguised variations in the units of measurements used, rather than being a genuine physical change (various claims that the speed of light 'c' may vary fall into this category [Ellis and Uzan, 2005]). This proposal has received impetus in recent times from ideas based in quantum field theory and string theory, suggesting that many of the 'constants of nature' are in fact contingent, depending on the nature of the vacuum state [Susskind, 2003; Freivogel *et al.*, 2005a]. This kind of proposal is to some degree open to observational test [Cowie and Songaila, 1995; Will, 1979], and in the cases where it has been investigated it seems that it does not occur in the visible region of the universe — the constants of nature are indeed invariant, with one possible exception: the fine structure constant, where there is claimed to be evidence of a very small change over astronomical timescales [Barrow, 2003]. That issue is still under investigation. Testing such invariance is fundamentally important, precisely because cosmology usually assumes as a ground rule that physics is the same everywhere in the universe. If this were not true, local physics would not guide us adequately as to the behaviour of matter elsewhere or at other times, and cosmology would become an arbitrary guessing game. In order to proceed in a scientific manner when such variation is proposed, one needs then to hypothesize the manner of such variation. Thus the old laws where G was constant are replaced by new laws governing its time variation [Brans and Dicke, 1961]; the principle of nature being governed by invariant (unchanging) physical laws and associated constants remains.[27] Thus in the end the proposal is to replace simpler old laws by new more complex ones. These must then be assumed invariant, or we cannot proceed scientifically.

(b) *Inertia and Mach's Principle*: It might be that the local inertial properties of matter are determined by the distant distribution of matter in the universe, so that if the universe were different, inertia would be different. This is the complex of ideas referred to as Mach's principle [Barbour and Pfister, 1995], which served as a major impetus for Einstein's cosmological ideas. The precise meaning and implications of this idea remain controversial.

(c) *The arrow of time*: The existence and direction of the macroscopic arrow of time in physics — and hence in chemistry, biology, psychology, and society — is related to boundary conditions in the past and future of the universe. The fundamental physical laws by themselves are time symmetric, and so unable to explain this feature [Davies, 1974; Ellis and Sciama, 1972; Zeh, 1992; Uffink, 2006]. A recent argument of this kind is Penrose's claim that the existence of the arrow of time is crucially based in the universe having had rather special initial conditions [Penrose, 1989b; Penrose, 1989a; Wald, 2005]. Thus what appears in ordinary physics as an immutable law of nature (viz. the Second Law of Thermodynamics with a given arrow of time) may well be the result of specific boundary conditions at the start and end of the universe. It might not be true in all universes, even if the underlying fundamental physical laws are the same.

In each case proposals have been made as to the possible nature of the deeper underlying unchanging laws, and the relations between the state of the universe and the resultant effective laws in that context. This is also proposed in the 'landscape' of possibilities of string theory [Susskind, 2005]. These proposals are however intrinsically untestable, for the reasons explained above (Sec. 3): we cannot change

[27] "Despite the incessant change and dynamic of the visible world, there are aspects of the fabric of the universe which are mysterious in their unshakeable constancy. It is these mysterious unchanging things that make our universe what it is and distinguish it from other worlds we might imagine" (Barrow [Barrow, 2003], p. 3).

the boundary conditions of the universe and see what happens; but they do serve as a continuing fertile source of ideas.

7.2 Alternative physics

In any case, the important conclusion is that it is certainly appropriate for cosmology to consider what would have happened if, not only the boundary conditions at the beginning of the universe, but also the laws of physics had been different [Susskind, 2005]:

Thesis E2: We cannot take the nature of the laws of physics for granted. *Cosmology is interested in investigating hypothetical universes where the laws of physics are different from those that obtain in the real universe in which we live — for this may help us understand why the laws of physics are as they are (a fundamental feature of the real physical universe).*

One cannot take the existence and nature of the laws of physics (and hence of chemistry) as unquestionable in cosmology — which seems to be the usual habit in biological discussions on the origin and evolution of life. This is in stark contrast to the rest of science, where we are content to take the existence and nature of the laws describing the fundamental behaviour of matter as given and unchangeable. Cosmological investigation is interested in the properties of hypothetical universes with different physical behaviour. Consideration of 'what might have been' is a useful cosmological speculation that may help throw light on what actually is; this is a statement of the usefulness of 'Gedanken experiments' in cosmology.

Indeed if one wants to investigate issues such as why life exists in the universe, consideration of this larger framework — in essence, a hypothetical ensemble of universes with many varied properties — is essential (this is of course not the same as assuming an ensemble of such universes actually exists, cf. the discussion below in Sec. 9.2). However we need to be very cautious about using any claimed statistics of universes in such a hypothetical ensemble of all possible or all conceivable universes. This is usually not well defined, and in any case is only relevant to physical processes if either the ensemble actually exists, rather than being a hypothetical one, or if it is the outcome of processes that produce well-defined probabilities — an untestable proposal. We can learn from such considerations the nature of possible alternatives, but not necessarily the probability with which they might occur (if that concept has any real meaning).

7.3 Emergence of complexity

As the universe evolves an increase of complexity takes place in local systems as new kinds of objects come into being that did not exist before — nuclei, atoms, stars and galaxies, planets, life, consciousness, and products of the mind such as books and computers [Morowitz, 2002]. New kinds of physical states come into being at late times such as Bose-Einstein condensates, that plausibly cannot exist without the intervention of intelligent beings.

Thesis E3: Physical novelty emerges in the expanding universe. *New kinds of physical existence come into being in the universe as it evolves, that did not exist previously. Their existence is allowed by the boundary conditions provided by the universe for local systems, together with the possibility space generated by the underlying physics. While their physical existence is novel, every new thing that comes into being is foreshadowed in possibility structures that precede their existence.*

Physical existence is new as the universe evolves, but there had to be precursors of the novel in the possibility space allowed by physics, so that they could come into being. In this sense the truly novel does not emerge *ex nihilo* but rather is discovered. The universe is the environment that allows this to happen. The nature of the features leading to the existence of life, and their possible causes, is discussed in Sec. 9.1.

8 ISSUE F: THE EXPLICIT PHILOSOPHICAL BASIS

Consequent on the discussion above, and particularly items **B6**, **C2**, and **D2**, it follows that

Thesis F1: Philosophical choices necessarily underly cosmological theory. *Unavoidable metaphysical issues inevitably arise in both observational and physical cosmology. Philosophical choices are needed in order to shape the theory.*

There is of course always a philosophical basis to any scientific analysis, namely adoption of the basic scientific method and a commitment to the attempt to explain what we see as far as possible simply in terms of causal laws, ultimately based in physics. This will clearly be true also in cosmology. However we need further explicit philosophical input in order to attain specific geometric models — for example a Copernican principle, as explained above, see Sec. 4.2.2 — and to determine what form physical cosmology should take in the very early universe, for example deciding which physical principle to use as the core of one's extrapolation of known physics to the unknown (Sec. 5). Underlying both sets of choices are criteria for satisfactoriness of a cosmological model, which help decide which feature to focus on in formulating a theory. Of particular importance is the scope chosen for our cosmological theory; together with the choice of criteria for a good theory, this is a philosophical decision that will shape the rest of the analysis. Some cosmologists tend to ignore the philosophical choices underlying their theories; but simplistic or unexamined philosophical standpoints are still philosophical standpoints!

8.1 Criteria for theories

As regards criteria for a good scientific theory [Kuhn, 1977], typical would be the following four areas of assessment:

1. *Satisfactory structure*: **(a)** internal consistency, **(b)** simplicity (Occam's razor), and **(c)** aesthetic appeal ('beauty' or 'elegance').

2. *Intrinsic explanatory power*: **(a)** logical tightness, **(b)** scope of the theory — the ability to unify otherwise separate phenomena, and **(c)** probability of the theory or model with respect to some well-defined measure;

3. *Extrinsic explanatory power, or relatedness*: **(a)** connectedness to the rest of science, **(b)** extendability — providing a basis for further development;

4. *Observational and experimental support*, in terms of **(a)** testability: the ability to make quantitative as well as qualitative predictions that can be tested; and **(b)** confirmation: the extent to which the theory is supported by such tests as have been made.

It is particularly the latter that characterizes a scientific theory, in contrast to other types of theories claiming to explain features of the universe and why things happen as they do. It should be noted that *these criteria are philosophical in nature in that they themselves cannot be proven to be correct by any experiment.* Rather their choice is based on past experience combined with philosophical reflection. One could attempt to formulate criteria for good criteria for scientific theories, but of course these too would need to be philosophically justified. The enterprise will end in infinite regress unless it is ended at some stage by a simple acceptance of a specific set of criteria.

Thesis F2: Criteria of satisfactoriness for theories cannot be scientifically chosen or validated. *Criteria of satisfactoriness are necessary for choosing good cosmological theories; these criteria have to be chosen on the basis of philosophical considerations. They should include criteria for satisfactory structure of the theory, intrinsic explanatory power, extrinsic explanatory power, and observational and experimental support.*

The suggestion here is that the above proposed criteria are a good set to use in investigating cosmology; they include those most typically used ([Kuhn, 1977]; and see [Penrose, 2004; Susskind, 2005] for comments on such criteria).

8.1.1 Conflicts between criteria.

These criteria are all acknowledged as desirable. The point then is that generally in pursuing historical sciences, and in particular in the cosmological context, they will not all be satisfied to the same degree, and may even lead to opposing conclusions:

Thesis F3: Conflicts will inevitably arise in applying criteria for satisfactory cosmological theories. *Philosophical criteria for satisfactory cosmological theories will in general come into conflict with each other, so that one will have to choose between them to some degree; this choice will shape the resulting theory.* [Ellis, 1991].

The thrust of much recent development has been away from observational tests towards strongly theoretically based proposals, indeed sometimes almost discounting observational tests. At present this is being corrected by a healthy move to detailed observational analysis of the consequences of the proposed theories, marking a maturity of the subject. However because of all the limitations in terms of observations and testing [criteria **(4)**], in the cosmological context we still have to rely heavily on other criteria, and some criteria that are important in most of science may not really make sense. This is true of **2(c)** in particular, as discussed above, see Sec. 3; nevertheless many approaches still give the idea of probability great weight. At a minimum, the ways this can make sense needs exploration and explication. Furthermore the meaning of some of the criteria may come into dispute. **1(b)** is clearly a case in point : for example, is the idea of an existent ensemble of universes displaying all possible behaviours simple (because it is a single idea that can be briefly stated), or immensely complex (because that statement hides all the complexities and ambiguities involved in the idea of an infinity of possibilities)? **1(c)** is also controversial('beauty is in the eye of the beholder'), see [Susskind, 2005] for a discussion.

The tenor of scientific understanding may change, altering the balance of what is considered a good explanation and what is not. An example [Ellis, 1990] is the way cosmologists strongly resisted the idea of an evolving universe in the 1920's, at a time when biological evolution was very well established but the idea of continental drift was also being strongly resisted. The change to an appreciation of the explanatory power of an evolving model came later in both cases; but even then in the cosmological case, for either aesthetic or metaphysical reasons, some still sought for a steady state description, resisting the implication of a beginning to the universe. That tendency is still with us today, in the form of models that are eternal in one way or another (e.g. some forms of chaotic inflation). Another example is the change from supposition of underlying order, expressed in the idea of a Cosmological Principle, to a broad supposition of generic disordered conditions, embodied in the ideas of inflation. Associated with this is a shift from making geometric assumptions to providing physical explanatory models. It is this shift that underlies the major present support for inflation:

Thesis F4: The physical reason for believing in inflation is its explanatory power as regards structure growth in the universe. *Inflation predicts the existence of Gaussian scale-free perturbations in the early universe thereby (given the presence of cold dark matter) explaining bottom-up structure formation in a satisfactory way. This theory has been vindicated spectacularly through observations of the CBR and matter power spectra. It is this explanatory power that makes it so acceptable to physicists, even though the underlying physics is neither well-defined nor tested, and its major large-scale observational predictions are untestable.*

The physical explanatory power of inflation in terms of structure formation, supported by the observational data on the fluctuation spectra, is spectacular. For most physicists, this trumps the lack of identification and experimental verification of the underlying physics (Sec. 5). Inflation provides a causal model that brings a wider range of phenomena into what can be explained by cosmology (Criterion **2(b)**), rather than just assuming the initial data had a specific restricted form. Explaining flatness ($\Omega_0 \simeq 1$ as predicted by inflation) and homogeneity reinforces the case, even though these are philosophical rather than physical problems (they do not contradict any physical law; things could just have been that way). However claims on the basis of this model as to what happens very far outside the visual horizon (as in the chaotic inflationary theory) results from prioritizing theory over the possibility of observational and experimental testing [Earman and Mosterin, 1999]. It will never be possible to *prove* these claims are correct.

8.2 The scope of cosmology

To sensibly choose priorities for the criteria just discussed, we need an answer to the question, How much should we try to explain?

Thesis F5: Cosmological theory can have a wide or narrow scope of enquiry. *The scope we envisage for our cosmological theory shapes the questions we seek to answer. The cosmological philosophical base becomes more or less dominant in shaping our theory according to the degree that we pursue a theory with more or less ambitious explanatory aims in terms of all of physics, geometry, and underlying fundamental causation.*

This is a choice one has to make, as regards both foundations and outcomes. Given a decision on this, one can sensibly debate what is the appropriate philosophical position to adopt in studying a cosmological theory with that scope. The study of expansion of the universe and structure formation from nucleosynthesis to the present day is essential and well-informed. The philosophical stance adapted is minimal and highly plausible. The understanding of physical processes at earlier times, back to quantum gravity, is less well founded. The philosophical stance is more significant and more debatable. Developments in the quantum gravity era are highly speculative; the philosophical position adopted is dominant because experimental and observational limits on the theory are lacking.

One can choose the degree to which one will pursue the study of origins [Fabian, 1989] back to earlier and earlier times and to more fundamental causal issues, and hence the degree to which specific philosophical choices are dominant in one's theory. The basic underlying cosmological questions are [Ellis, 1991]:

1. *Why do the laws of physics have the form they do?* Issues arise such as what makes particular laws work? For example, what guarantees the behaviour of a proton, the pull of gravity? What makes one set of physical laws 'fly' rather than another? If for example one bases a theory of cosmology on string theory [Susskind, 2005], then who or what decided that quantum gravity would have a nature well described by string theory? If one considers

all possibilities, considering string theory alone amounts to a considerable restriction.

2. *Why do boundary conditions have the form they do ?* The key point here (Sec. 6.2), is how are specific contingent choices made between the various possibilities, for example whether there was an origin to the universe or not.

3. *Why do any laws of physics at all exist ?* This relates to unsolved issues concerning the nature of the laws of physics: are they descriptive or prescriptive? (Sec. 9.3.3). Is the nature of matter really mathematically based in some sense, or does it just happen that its behaviour can be described in a mathematical way?

4. *Why does anything exist ?* This profound existential question is a mystery whatever approach we take.[28]

   Finally the adventurous also include in these questions the more profound forms of the contentious Anthropic question [Carr and Rees, 1979; Davies, 1982; Barrow and Tipler, 1984; Tegmark, 1998; Susskind, 2005]:

5. *Why does the universe allow the existence of intelligent life?* This is of somewhat different character than the others and largely rests on them but is important enough to generate considerable debate in its own right.

The status of all these questions is philosophical rather than scientific, for they cannot be resolved purely scientifically. How many of them — if any — should we consider in our construction of and assessments of cosmological theories?

One option is *to decide to treat cosmology in a strictly scientific way*, excluding all the above questions, because they cannot be solved scientifically. One ends up with a solid technical subject that by definition excludes such philosophical issues. This is a consistent and logically viable option. This logically unassailable position however has little explanatory power; thus most tend to reject it because of criteria **2(b)** and **3** above.

The second option is to decide that *these questions are of such interest and importance that one will tackle some or all of them, even if that leads one outside the strictly scientific arena.* It is here that criteria **2** and **3** above are to some degree in conflict with criterion **4**. Thus if we try to explain the origin of the universe itself, these philosophical choices become dominant precisely because the experimental and observational limits on the theory are weak; this can be seen by viewing the variety of such proposals that are at present on the market.

8.3 Limits of Representation and Knowledge of Reality

It follows from the above discussion that there are limits to what the scientific method can achieve in explanatory terms. We need to respect these limits and

[28] But see Grunbaum [Grunbaum, 2004] for a dissenting view.

acknowledge clearly when arguments and conclusions are based on some philosophical stance rather than purely on testable scientific argument. If we acknowledge this and make that stance explicit, then the bases for different viewpoints are clear and alternatives can be argued about rationally.

A crucial underlying feature here is relating the nature of epistemology to ontology: how do we relate evidence to our theories of existence? A further key issue is the relation of models to reality:

Thesis F6: Reality is not fully reflected in either observations or theoretical models. *Problems arise from confusion of epistemology (the theory of knowledge) with ontology (the nature of existence): existence is not always manifest clearly in the available evidence. The theories and models of reality we use as our basis for understanding are necessarily partial and incomplete reflections of the true nature of reality, helpful in many ways but also inevitably misleading in others. They should not be confused with reality itself!*

The confusion of epistemology with ontology occurs all the time, underlying for example the errors of both logical positivism and extreme relativism. In particular, it is erroneous to assume that lack of evidence for the existence of some entity is proof of its non-existence. In cosmology it is clear for example that regions may exist from which we can obtain no evidence (because of the existence of horizons); so we can sometimes reasonably deduce the existence of unseen matter or regions from a sound extrapolation of available evidence (no one believes matter ends at or just beyond the visual horizon). However one must be cautious about the other extreme, assuming existence can always be assumed because some theory says so, regardless of whether there is any evidence of existence or not. This happens in present day cosmology, for example in presentations of the case for multiverses, even though the underlying physics has not been experimentally confirmed. It may be suggested that arguments ignoring the need for experimental/observational verification of theories ultimately arise because these theories are being confused with reality, or at least are being taken as completely reliable total representations of reality. This occurs in

- Confusing computer simulations of reality with reality itself, when they can in fact represent only a highly simplified and stylized version of what actually is;
- Confusing the laws of physics themselves with their abstract mathematical representation (if indeed they are ontologically real, c.f. Sec. 10.1), or confusing a construction of the human mind ('Laws of Physics') with the reliable behaviour of ponderable matter (if they are not ontologically real);
- Confusing theoretically based outcomes of models with proven observational results (e.g. claiming the universe necessarily has flat spatial sections: $\Omega_0 = 1$, and so this can be taken for granted, when the value of Ω_0 can and should be observationally determined precisely because this then tests that prediction).

No model (literary, intuitive, or scientific) can give a perfect reflection of reality. Such models are always selective in what they represent and partial in the completeness with which they do so. The only model that would reflect reality fully is a perfect fully detailed replica of reality itself! This understanding of the limits of models and theories does not diminish the utility of these models; rather it helps us use them in the proper way. This is particularly relevant when we consider how laws of nature may relate to the origins of the universe itself, and to the existence and nature of life in the expanding universe. The tendency to rely completely on our theories, even when untested, seems sometimes to arise because we believe they are the same as reality — when at most they are *descriptions* of reality.

9 KEY ISSUES

There are some interrelated key issues where the features identified above either are at the heart of current debates, or are likely to be at the heart of future debates. They are: the reason cosmological conditions allow the existence of life (anthropic issues), the closely related issue of the possible existence of multiverses; and the natures of existence, including the questions of the existence of infinities and the nature of the laws of physics. We look at them in turn in this section. To some degree they have already been considered above, but they are specifically featured here because of the important role they will probably play in discussion in the future.

9.1 Issue G: The anthropic question: Fine tuning for life

One of the most profound fundamental issues in cosmology is the Anthropic question, see [Davies, 1982; Barrow and Tipler, 1984; Earman, 1987; Fabian, 1989; Davies, 1987; Balashov, 1991; Rees, 1999; Rees, 2003; Barrow, 2003]: *why does the Universe have the very special nature required in order that life can exist?*. The point is that a great deal of "fine tuning" is required in order that life be possible. There are many relationships embedded in physical laws that are not explained by physics, but are required for life to be possible; in particular various fundamental constants are highly constrained in their values if life as we know it is to exist:
"A universe hospitable to life — what we might call a biophilic universe — has to be special in many ways ... Many recipes would lead to stillborn universes with no atoms, no chemistry, and no planets; or to universes too short lived or too empty to evolve beyond sterile uniformity" [Rees, 2003].
How has it come about that the Universe permits the evolution and existence of intelligent beings at any time or place? "What features of the universe were essential for creatures such as ourselves, and is it through coincidence or for some deeper reason that our universe has these features?" [Gribbin and Rees, 1991]. Whether one regards this as an appropriate issue for cosmology to discuss depends, as discussed above (Sec. 8.2), on the scope one envisages for cosmology. The viewpoint taken here will be that this is one of the major issues one might wish

to explain, and indeed a substantial literature considers this. Here we explore the nature of this fine tuning, and then consider possible answers as to how it arises. There are three aspects that we consider in turn (cf. [Susskind, 2005]).

9.1.1 Laws of physics and the existence of complexity

The laws of physics and chemistry are such as to allow the functioning of living cells, individuals, and ecosystems of incredible complexity and variety, and it is this that has made evolution possible. What requires explanation, *is why the laws of physics are such as to allow this complex functionality to work*, without which no evolution whatever would occur. We can conceive of universes where the laws of physics (and so of chemistry) were different than in ours. Almost any change in these laws will prevent life as know it from functioning.

The first requirement is *the existence of laws of physics that guarantee the kind of regularities that can underlie the existence of life.* These laws as we know them are based on variational and symmetry principles; we do not know if other kinds of laws could produce complexity. If the laws are in broad terms what we presently take them to be, the following *inter alia* need to be right, for life of the general kind we know to exist [Davies, 1982; Gribbin and Rees, 1991]:

- Quantization that stabilizes matter and allows chemistry to exist through the Pauli exclusion principle.
- The neutron-proton mass differential must be highly constrained. If the neutron mass were just a little less than it is, proton decay could have taken place so that by now no atoms would be left at all [Davies, 1982].
- Electron-proton charge equality is required to prevent massive electrostatic forces overwhelming the weaker electromagnetic forces that govern chemistry.
- The strong nuclear force must be strong enough that stable nuclei exist [Davies, 1982]; indeed complex matter exists only if the properties of the nuclear strong force lies in a tightly constrained domain relative to the electromagnetic force [Tegmark, 2003].
- The chemistry on which the human body depends involves intricate folding and bonding patterns that would be destroyed if the fine structure constant (which controls the nature of chemical bonding) were a little bit different.
- The number D of large spatial dimensions must be just 3 for complexity to exist [Tegmark, 2003; Rees, 2003].

Hogan has examined the freedom in the parameters of the standard model of particle physics and concluded that 5 of the 17 free parameters of the standard model must lie in a highly constrained domain if complex structures are to exist [Hogan, 2003]. This is of course taking the basic nature of the standard model of particle physics for granted. If this were not so, it is difficult to determine what the constraints would be. However his study is sufficient to show that whatever the nature of fundamental physics, and in particular of particle physics, may be, only a small subset of all possible laws of physics will be compatible with the existence of complexity.

9.1.2 Laws of physics and the existence of congenial environments

The creation through astrophysical processes of suitable habitats for life to exist (the existence of planets circling stable stars, for example) depends to some degree on the nature of the fundamental physical laws. If the laws are in broad terms what we presently take them to be, the requirements for such habitats to exist include:

- The gravitational force must create large stable structures (planets and stars) that can be the habitat for life and their energy source respectively. This requires the gravitational force to be very weak relative to electrical forces. The ratio $\mathcal{N}$ of the strength of the electromagnetic force to the gravitational force must be close to the observed value: $\mathcal{N} \simeq 10^{36}$ [Rees, 1999, Ch. 3].
- The weak force must allow helium production that leaves sufficient hydrogen over; it is related to gravity through a numerical factor of 10^{-11}, which cannot be much different. And for this to work, the neutron-proton mass difference must be close to the mass of the electron [Davies, 1982].
- A stellar balance should allow a long lifetime for stars like the sun, so allowing the transmutation of the light elements into heavy elements. This requires that the nuclear fusion efficiency $\mathcal{E}$ be close to the observed value: $\mathcal{E} \simeq 0.007$ [Rees, 1999, Ch. 4].
- One needs to overcome the beryllium "bottleneck" in the making of heavy elements through nuclear reactions in stars [Gribbin and Rees, 1991; Susskind, 2005]. The production of carbon and oxygen in stars requires the careful setting of two different nuclear energy levels to provide a resonance; if these levels were just a little different, the elements we need for life would not exist [Fabian, 1989]. Indeed it was on this basis that Hoyle famously predicted a carbon-12 energy level that has since been experimentally confirmed.
- One needs something like the existence of neutrinos and the weak interaction with its specific coupling constant in order to underly supernovae explosions that spread heavy elements through space, as seeds for planetary formation [Gribbin and Rees, 1991].
- The nuclear force must be weak enough that di-protons do not exist, otherwise no protons will be left over to enable heavier elements to exist [Davies, 1982].
- The neutrino mass must not be too high, or the universe will not last long enough [Davies, 1982].

9.1.3 Cosmological boundary/initial conditions and congenial environments

Finally, given laws of physics that are suitable in terms of satisfying the requirements of both the previous sections, the universe itself must also be suitable, in terms of its initial or boundary conditions, for life to exist. If the laws of physics are basically the same as we now believe them to be, these cosmological requirements include

- The size of the universe and its age must be large enough. There could be universes that expanded and then recollapsed with a total lifetime of only 100,000 years; we need a sufficiently old universe for second generation stars to come into existence and then for planets to have a stable life for long enough that evolution could lead to the emergence of intelligent life. Thus the universe must be at about 15 billion years old for life to exist [Gribbin and Rees, 1991], hence we must have $\Omega_{matter} \simeq 0.3$ [Rees, 1999, Ch. 6].

- The size of the cosmological constant must not be too large, or galaxies will not form; we need $|\Omega_\Lambda| < 1$ for galaxies to exist [Rees, 1999, Ch. 7]; [Susskind, 2005].

- The seeds in the early universe for fluctuations that will later grow into galaxies must be of the right size that structures form without collapsing into black holes: the number Q characterizing the size of primordial ripples on the LSS (and hence the geometry of the perturbed cosmological model, see Sec. 2.5.2) must therefore be of the order $Q \simeq 10^{-5}$ [Rees, 1999, Ch. 8].

The complex of interacting systems in a human body could not possibly work if a series of delicate conditions were not maintained. For example, the background radiation might never drop below 3000 K, so that matter was always ionized (electrons and nuclei always remaining separate from each other); the molecules of life could then never form. Black holes might be so common that they rapidly attracted all the matter in the universe, and there never was a stable environment in which life could develop. Cosmic rays could always be so abundant that any tentative organic structures were destroyed before they could replicate. Overall,

- There must be non-interference with local systems. The concept of locality is fundamental, allowing local systems to function effectively independently of the detailed structure of the rest of the Universe. We need the universe and the galaxies in it to be largely empty, and gravitational waves and tidal forces to be weak enough,[29] so that local systems can function in a largely isolated way [Ellis, 2002].

- The fact that the night sky is dark ('Olbers' paradox' [Bondi, 1960; Harrison, 2000]) is a consequence of the expansion of the universe together with the photon to baryon ratio. This feature is a necessary condition for the existence

[29]Thus the Weyl tensor C_{abcd} must be suitably small everywhere, presumably implying an almost-RW geometry, cf. [Stoeger *et al.*, 1995].

of life: the biosphere on Earth functions by disposing of waste energy to the heat sink of the dark night sky [Penrose, 1989b]. Thus one way of explaining why the sky is observed to be dark at night is that if this were not so, we would not be here to observe it.

- The existence of the arrow of time, and hence of laws like the second law of thermodynamics, are probably necessary for evolution and for consciousness. This depends on boundary conditions at the beginning and end of the Universe (Sec. 7.1).
- Presumably the emergence of a classical era out of a quantum state is required. The very early universe would be a domain where quantum physics would dominate, leading to complete uncertainty and an inability to predict the consequence of any initial situation; we need this to evolve to a state where classical physics leads to the properties of regularity and predictability that allow order to emerge.
- Physical conditions on planets must be in a quasi-equilibrium state for long enough to allow the delicate balances that enable our existence, through the very slow process of evolution, to be fulfilled.

Thus the existence of suitable local systems to be a habitat for life depends critically on the large-scale properties of very distant matter. These provides a stable local environment within which life can develop.

9.1.4 Fine tuning overall

Thus there are many ways that conditions in a universe could prevent life occurring. Life will occur only if: there exist heavy elements; there is sufficient time for evolution of advanced life forms to take place; there are regions in the universe that are neither too hot nor too cold; there are precisely restricted values of the fundamental constants that control chemistry and local physics; and so on. These conditions will not be true in a generic universe. In summary,

Thesis G1: Life is possible because both the laws of physics and the boundary conditions for the universe have a very special nature. *Only particular laws of physics, and particular initial conditions in the Universe, allow the existence of intelligent life of the kind we know. No evolutionary process whatever is possible for any kind of life if these laws and conditions do not have this restricted form.*

Why is this so? One should note that we can only meaningfully refer here to 'life as we know it'. One of the recurring issues is whether there could be some other quite different basis for life. You can if you wish speculate that life might exist in some immaterial form, or based only on light elements, or existent deep in empty space without the need for stars or planets to provide a viable habitat. The anthropic literature is based on assuming this is not viable, but we cannot *prove* anything in this regard. We have no idea of any basis by which life might come into existence other than the broad principles we see in the life around us. The basic principles of life as we understand it require a great degree of complex organization enabling it to fulfil a complex variety of functions that can only, as far as we know, be based in material existence with information storage, energy usage, sensing of the external world, etc., which requires at a minimum heavy elements (carbon, nitrogen, oxygen, phosphorus for example), a long-term energy source (such as the flow of energy from the sun), and a stable environment (such as the surface of a planet). When we abandon this basis for understanding — saying ‘yes but some other form of life might exist’ without providing any proposal for its possible structure — one enters the unprofitable realm of speculation. It does not seem to provide any useful way forward.

9.1.4.1 The Weak Anthropic Principle. There are two purely scientific approaches to the Anthropic issue.[30] The first is the *Weak Anthropic Principle* (WAP), based on the comment: it is not surprising the observed Universe admits the existence of life, for the Universe cannot be observed unless there are observers in it [Barrow and Tipler, 1984; Balashov, 1991]. This seemingly empty statement gains content when we turn it round and ask, at what times and places in the Universe can life exist, and what are the inter-connections that are critical for its existence? It could not for example exist too early in the present expansion phase, for the night sky would then have been too hot. Furthermore one can deduce various necessary relations between fundamental quantities in order that the observers should exist (e.g. those mentioned above), so that if for example the fundamental constants vary with time or place in the Universe, life will only be possible in restricted regions where they take appropriate Anthropic values.

Hence this view basically interprets the Anthropic principle as a selection principle: the necessary conditions for observers to exist restricts the times and places from which the Universe can be observed. Because it is quite possible that conditions would not be right for life to exist anywhere in an arbitrarily selected universe, it is also usually conjoined with the idea of the existence of a multiverse, as discussed below, see Sec. 9.2. This is an interesting and often illuminating viewpoint. For example, neither the Chaotic Inflationary Universe idea (Sec. 2.6) nor any other multiverse proposal works unless we add such an Anthropic component into their interpretation to explain why we observe the Universe from a viewpoint where life exists. It is now used by some physicists to explain the low value of the cosmological constant (which quantum field theory predicts should have a very much larger value than observed, see Sec. 9.2.5), and occurs in the context of the possibility landscape of string theory [Susskind, 2005].

[30] I omit the so-called *Final Anthropic Principle* (FAP for short), which maintains that intelligent life must necessarily evolve and then remain in existence until the end of the universe, for I do not believe it merits serious discussion as a scientific proposal; indeed it led to a famous book review referring to the *Completely Ridiculous Anthropic Principle* (CRAP for short) [Gardner, 1986].

9.1.4.2 The Strong Anthropic Principle. By contrast, the *Strong Anthropic Principle* (SAP) [Barrow and Tipler, 1984; Balashov, 1991] claims that it is necessary that intelligent life exist in the Universe; the presence of life is required in order that a universe model make sense. This is clearly a very controversial claim, for it is hard to provide scientific reasons to support this view. One can suggest that the most solid justification attempted is through the claim that the existence of an observer is necessary in order that quantum theory can make sense. However, this justification is based on one of a number of different interpretations of quantum theory; the nature of these quantum foundations is controversial, and not resolved [Isham, 1997; Dickson, 2006; Landsman, 2006].

Furthermore if we were to suppose this justification correct, then the next step is to ask: Why does the Universe need quantum mechanics anyway? The argument would be complete only if we could prove that quantum mechanics was absolutely necessary for every self-consistent Universe; but that line of reasoning cannot be completed at present, not least because quantum mechanics itself is not a fully self-consistent theory. Apart from the conceptual problems at its foundation due to the unresolved measurement issue [Isham, 1997], it suffers from divergences that so far have proved irremediable in the sense that we can work our way round them to calculate what we need, but cannot remove them. The SAP proposal has no accepted physical foundation, and also raises problematic philosophical issues [Earman, 1987]. I will not pursue it further here.

9.1.5 The relation to fundamental physical theories

Many physicists go further, rejecting any Anthropic form of reasoning. They regard it as a cop-out resorted to when physical theories fail to give the needed answers, and seek to obtain a full answer from physics alone [Scott, 2005; Susskind, 2005]. One possibility is that there is a fundamental theory of everything that determines the nature of physics completely, with no arbitrary parameters left, and this still to be discovered theory just happens to be of such a nature as to admit life.

However in this case the Anthropic issue returns with a vengeance: *How could it be that such a theory, based for example on variational principles and the specific invariance groups of particle physics, could just happen to lead to biophilic parameter values*? There is no clear way to answer such a question. Uniqueness of fundamental physics resolves the parameter freedom only at the expense of creating an even deeper mystery, with no way of resolution apparent. In effect, the nature of the unified fundamental force would be pre-ordained to allow, or even encourage, the existence of life; but there would be no apparent reason why this should be so.

A second possibility is that physics allows many effective theories with varying parameters — some form of multiverse, as for example may be implied by string theory [Susskind, 2003; Freivogel *et al.*, 2005a; Susskind, 2005]. If these varying options are all equally real, life can occur because in some cases the parameters will lie in the restricted biophilic regime. Thus from this viewpoint the Anthropic idea is intimately linked with the existence of multiverses, which provide a legitimate domain for their application. We will turn to an examination of multiverses in the next section, but before doing so we will consider the range of metaphysical options for resolving the anthropic question.

9.1.6 The metaphysical options

To make progress on the Anthropic issue, we have to seriously consider the nature of ultimate causation: What is the fundamental cause for the phenomena we see? If we pursue the chain of physical cause and effect to its conclusion, we are still left with the question: *Why did this occur, and not something else*? Whatever the reason is, it is the ultimate cause we are seeking. Note that we are here leaving the terrain of science itself, and starting to probe the domain of metaphysics — the foundations of science and indeed of existence. As noted above, one can simply decide not to pursue such issues. If we do continue to question, there appear to be basically six approaches to the issue of ultimate causation: namely Random Chance, Necessity, High Probability, Universality, Cosmological Natural Selection, and Design. We briefly consider these in turn.

Option 1: *Random Chance, signifying nothing.* The initial conditions in the Universe just happened, and led to things being the way they are now, by pure chance. Probability does not apply. There is no further level of explanation that applies; searching for 'ultimate causes' has no meaning.

This is certainly logically possible, but not satisfying as an explanation, as we obtain no unification of ideas or predictive power from this approach. Nevertheless some implicitly or explicitly hold this view.

Option 2: *Necessity.* Things have to be the way they are; there is no other option. The features we see and the laws underlying them are demanded by the unity of the Universe: coherence and consistency require that things must be the way they are; the apparent alternatives are illusory. Only one kind of physics is self-consistent: all logically possible universes must obey the same physics.

To really prove this would be a very powerful argument, potentially leading to a self-consistent and complete scientific view. But we can imagine alternative universes! — why are they excluded? Furthermore we run here into the problem that we have not succeeded in devising a fully self-consistent view of physics: neither the foundations of quantum physics nor of mathematics are on a really solid consistent basis. Until these issues are resolved, this line cannot be pursued to a successful conclusion.

Option 3: *High probability.* Although the structure of the Universe appears very improbable, for physical reasons it is in fact highly probable.

These arguments are only partially successful, even in their own terms. They run into problems if we consider the full set of possibilities: discussions proposing this kind of view actually implicitly or explicitly restrict the considered possibilities *a priori*, for otherwise it is not very likely the Universe will be as we see it. Besides, we do not have a proper measure to apply to the set of initial conditions, enabling us to assess these probabilities. Furthermore, as discussed above, see Sec. 3, application of probability arguments to the Universe itself is dubious, because the Universe is unique. Despite these problems, this approach has considerable support in the scientific community, for example it underlies the chaotic inflationary proposal (Sec. 2.6). It attains its greatest power in the context of the assumption of universality:

Option 4: *Universality.* This is the stand that "All that is possible, happens": an ensemble of universes or of disjoint expanding universe domains is realized in reality, in which all possibilities occur [Rees, 1999; Rees, 2003; Tegmark, 2003]. In its full version, the anthropic principle is realized in both its strong form (if all that is possible happens, then life must happen) and its weak form (life will only occur in some of the possibilities that are realized; these are picked out from the others by the WAP, viewed as a selection principle). There are four ways this has been pursued.

1. *Spatial variation.* The variety of expanding universe domains is realised in space through random initial conditions, as in chaotic inflation (Sec. 2.6). While this provides a legitimate framework for application of probability, from the viewpoint of ultimate explanation it does not really succeed, for there is still then one unique Universe whose (random) initial conditions need explanation. Initial conditions might be globally statistically homogeneous, but also there could be global gradients in some physical quantities so that the Universe is not statistically homogeneous; and these conditions might be restricted to some domain that does not allow life. It is a partial implementation of the ensemble idea; insofar as it works, it is really a variant of the "high probability" idea mentioned above. If it was the more or less unique outcome of proven physics, then that would provide a good justification; but the physics underlying such proposals is not even uniquely defined, much less tested. Simply claiming a particular scalar field with some specific stated potential exists does not prove that it exists!

2. *Time variation.* The variety of expanding universe domains could be realised across time, in a universe that has many expansion phases (a Phoenix universe), whether this occurs globally or locally. Much the same comments apply as in the previous case.

3. *Quantum Mechanical.* It could occur through the existence of the Everett-Wheeler "many worlds" of quantum cosmology, where all possibilities occur through quantum branching [Deutsch, 1998]. This is one of the few genuine alternatives proposed to the Copenhagen interpretation of quantum mechanics, which leads to the necessity of an observer, and so potentially to the

Strong Anthropic interpretation considered above (see Sec. 9.1). The many-worlds proposal is controversial: it occurs in a variety of competing formulations [Isham, 1997], none of which has attained universal acceptance. The proposal does not provide a causal explanation for the particular events that actually occur: if we hold to it, we then have to still explain the properties of the particular history we observe (for example, why does our macroscopic universe have high symmetries when almost all the branchings will not?). And above all it is apparently untestable: there is no way to experimentally prove the existence of all those other branching universes, precisely because the theory gives the same observable predictions as the standard theory.

4. *Completely disconnected*. They could occur as completely disconnected universes: there really is an ensemble of universes in which all possibilities occur, without any connection with each other [Lewis, 1986; Rees, 2003; Tegmark, 2003]. A problem that arises then is, What determines what is possible? For example, what about the laws of logic themselves? Are they inviolable in considering all possibilities? We cannot answer, for we have no access to this multitude of postulated worlds. We explore this further below (Sec. 9.2).

In all these cases, major problems arise in relating this view to testability and so we have to query the meaningfulness of the proposals as scientific explanations. They all contradict Occam's razor: we "solve" one issue at the expense of envisaging an enormously more complex existential reality. Furthermore, they do not solve the ultimate question: *Why does this ensemble of universes exist*? One might suggest that ultimate explanation of such a reality is even more problematic than in the case of single universe. Nevertheless this approach has an internal logic of its own which some find compelling. We consider this approach further below, see Sec. 9.2.

Option 5: *Cosmological Natural Selection*. If a process of re-expansion after collapse to a black hole were properly established, it opens the way to the concept not merely of evolution of the Universe in the sense that its structure and contents develop in time, but in the sense that the Darwinian selection of expanding universe regions could take place, as proposed by Smolin [Smolin, 1992]. The idea is that there could be collapse to black holes followed by re-expansion, but with an alteration of the constants of physics through each transition, so that each time there is an expansion phase, the action of physics is a bit different. The crucial point then is that some values of the constants will lead to production of more black holes, while some will result in less. This allows for evolutionary selection favouring the expanding universe regions that produce more black holes (because of the favourable values of physical constants operative in those regions), for they will have more "daughter" expanding universe regions. Thus one can envisage natural selection favouring those physical constants that produce the maximum number of black holes.

The problem here is twofold. First, the supposed 'bounce' mechanism has never been fully explicated. Second, it is not clear — assuming this proposed process can be explicated in detail — that the physics which maximizes black hole production is necessarily also the physics that favours the existence of life. If this argument could be made water-tight, this would become probably the most powerful of the multiverse proposals.

Option 6: *Purpose or Design.* The symmetries and delicate balances we observe require an extraordinary coherence of conditions and cooperation of causes and effects, suggesting that in some sense they have been purposefully designed. That is, they give evidence of intention, both in the setting of the laws of physics and in the choice of boundary conditions for the Universe. This is the sort of view that underlies Judaeo-Christian theology. Unlike all the others, it introduces an element of meaning, of signifying something. In all the other options, life exists by accident; as a chance by-product of processes blindly at work.

The prime disadvantage of this view, from the scientific viewpoint, is its lack of testable scientific consequences ("Because God exists, I predict that the density of matter in the Universe should be x and the fine structure constant should be y"). This is one of the reasons scientists generally try to avoid this approach. There will be some who will reject this possibility out of hand, as meaningless or as unworthy of consideration. However it is certainly logically possible. The modern version, consistent with all the scientific discussion preceding, would see some kind of purpose underlying the existence and specific nature of the laws of physics and the boundary conditions for the Universe, in such a way that life (and eventually humanity) would then come into existence through the operation of those laws, then leading to the development of specific classes of animals through the process of evolution as evidenced in the historical record. Given an acceptance of evolutionary development, it is precisely in the choice and implementation of particular physical laws and initial conditions, allowing such development, that the profound creative activity takes place; and this is where one might conceive of design taking place.[31]

However from the viewpoint of the physical sciences *per se*, there is no reason to accept this argument. Indeed from this viewpoint there is really no difference between design and chance, for they have not been shown to lead to different physical predictions.

9.1.7 Metaphysical Uncertainty

In considering ultimate causation underlying the anthropic question, in the end we are faced with a choice between one of the options above. As pointed out already by Kant and Hume, although we may be able to argue strongly for one or other of them, we cannot *prove* any of the options are correct [Hume, 1993].

[31] This is not the same as the view proposed by the 'Intelligent Design' movement. It does not propose that God tweaks the outcome of evolutionary processes.

Thesis G2: Metaphysical uncertainty remains about ultimate causation in cosmology. *We cannot attain certainty on the underlying metaphysical cosmological issues through either science or philosophy.*

If we look at the anthropic question from a purely scientific basis, we end up without any resolution, basically because science attains reasonable certainty by limiting its considerations to restricted aspects of reality; even if it occasionally strays into the area of ultimate causation, it is not designed to deal with it. By itself, it cannot make a choice between these options; there is no relevant experiment or set of observations that can conclusively solve the issue. Thus a broader viewpoint is required to make progress, taking into account both the scientific and broader considerations. The issue is of a philosophical rather than scientific nature. One important issue that then arises is what kind of data is relevant to these philosophical choices, in addition to that which can be characterized as purely scientific data (Sec. 9.3.4).

9.2 Issue H: The possible existence of multiverses

If there is a large enough ensemble of numerous universes with varying properties, it may be claimed that it becomes virtually certain that some of them will just happen to get things right, so that life can exist; and this can help explain the fine-tuned nature of many parameters whose values are otherwise unconstrained by physics [Rees, 1999; Rees, 2003]. As discussed in the previous section, there are a number of ways in which, theoretically, multiverses could be realized [Lewis, 1986; Tegmark, 2003]. They provide a way of applying probability to the universe [Sciama, 1971; Bostrom, 2002] (because they deny the uniqueness of the universe). However, there are number of problems with this concept. Besides, this proposal is observationally and experimentally untestable; thus its scientific status is debatable.

9.2.1 Definition

In justifying multiverses, it is often stated that 'all that can occur, occurs' (or similarly). However that statement does not adequately specify a multiverse. To define a multiverse properly requires two steps [Ellis *et al.*, 2004]. First, one needs to specify what is conceived of in the multiverse, by defining a *possibility space*: a space $\mathcal{M}$ of all possible universes, each of which can be described in terms of a set of states s in a state space $\mathcal{S}$. Each universe m in $\mathcal{M}$ will be characterized by a set of distinguishing parameters p, which are coordinates on $\mathcal{S}$. Choices are needed here. In geometrical terms, will it include only Robertson–Walker models, or more general ones (e.g. Bianchi models, or models without symmetries)? In gravitational terms, will it include only General Relativity, or also brane theories, models with varying G, loop quantum gravity models, string theory models with their associated possibility 'landscapes', and models based on the wave function of the universe concept? Will it allow only standard physics but with varying constants, or a much wider spectrum of physical possibilities, e.g. universes without quantum theory, some with five fundamental forces instead of four, and others with Newtonian gravity? Defining the possibility space means making some kind of assumptions about physics and geometry that will then apply across the whole family of models considered possible in the multiverse, and excluding all other possibilities.

Second, one needs to specify which of the possible universes are physically realized in the multiverse, and how many times each one occurs. *A multiverse must be a physically realized multiverse and not a hypothetical or conceptual one if it is to have genuine explanatory power.* Thus one needs a distribution function $f(m)$ specifying how many times each type of possible universe m in $\mathcal{M}$ is realised. The function $f(m)$ expresses the contingency in any actualization. Things could have been different! Thus, $f(m)$ describes a specific *ensemble of universes* or *multiverse* envisaged as being realised out of the set of possibilities. For example, $f(m)$ might be non-zero for all possible values of all the parameters p ('all that can happen, happens'); but it could be that f describes a multiverse where there are 10^{100} identical copies of one particular universe (the realization process finds a particularly successful recipe, and then endlessly replicates it).

Additionally we need a measure $d\pi$ that enables this function to determine numbers and probabilities of various properties in the multiverse: the number of universes corresponding to a set of parameter increments will be $\mathrm{d}N$ given by

$$(38)\quad \mathrm{d}N = f(m)d\pi$$

for continuous parameters; for discrete parameters, we add in the contribution from all allowed parameter values. The total number of universes N in the ensemble will be given by

$$(39)\quad N = \int_{\mathcal{M}} f(m)d\pi$$

(which will often diverge), where the integral ranges over all allowed values of the member parameters and we take it to include all relevant discrete summations. The expectation value P of a quantity $p(m)$ defined on the set of universes will be given by

$$(40)\quad P = \int_{\mathcal{M}} p(m)f(m)d\pi.$$

These three elements (the possibility space, the measure, and the distribution function) must all be clearly defined in order to give a proper specification of a multiverse [Ellis *et al.*, 2004]. This is almost never done.

9.2.2 Non-uniqueness: Possibilities

There is non-uniqueness at both steps. Stating "all that is possible, happens" does not resolve what is possible. The concept of multiverses is not well defined until the space of possible universes has been fully characterized; it is quite unclear how to do this uniquely. The issue of what is to be regarded as an ensemble of 'all possible' universes can be manipulated to produce any result you want, by redefining what is meant by this phrase — standard physics and logic have no necessary sway over them: what I envisage as 'possible' in such an ensemble may be denied by you. What super-ordinate principles are in operation to control the possibilities in the multiverse, and why? A key point here is that *our understandings of the possibilities are always of necessity arrived at by extrapolation from what we know,* and my imagination may be more fertile than yours, and neither need correspond to what really exists out there — if indeed there is anything there at all. Do we include only

- *Weak variation*: e.g. only the values of the constants of physics are allowed to vary? This is an interesting exercise but is certainly not an implementation of the idea 'all that can happen, happens'. It is an extremely constrained set of variations.

- *Moderate variation*: different symmetry groups, or numbers of dimensions, etc. We might for example consider the possibility landscapes of string theory [Freivogel *et al.*, 2005b] as realistic indications of what may rule multiverses [Susskind, 2003; Freivogel *et al.*, 2005a; Susskind, 2005]. But that is very far indeed from 'all that is possible', for that should certainly include spacetimes not ruled by string theory.

- *Strong variation*: different numbers and kinds of forces, universes without quantum theory or in which relativity is untrue (e.g. there is an aether), some in which string theory is a good theory for quantum gravity and others where it is not, some with quite different bases for the laws of physics (e.g. no variational principles).

- *Extreme variation*: universes where physics is not well described by mathematics; with different logic; universes ruled by local deities; allowing magic as in the Harry Potter series of books; with no laws of physics at all? Without even mathematics or logic?

Which is claimed to be the properties of the multiverse, and why? We can express our dilemma here through the paradoxical question: *Are the laws of logic necessary in all possible universes?*

9.2.3 Non-uniqueness: existence and causation

A specific multiverse is defined by specifying the distribution function $f(m)$ of actually realized universes. It is unclear what mechanism can underlie such a distribution, and any proposal for such a mechanism is completely untestable. We need some indication as to *what determines existence within the possibilities defined by the supposed possibility space*: What decides how many times each one happens?

Unless we understood the supposed underlying mechanisms we can give no serious answer; and there is no prospect whatever of testing any proposed mechanism. The mechanisms supposed to underlie whatever regularities there are in the multiverse must pre-exist the existence of not merely this universe but also every other one. If one assumes a universe that is connected in the large but is locally separated into causally disconnected domains with different physical properties(as in chaotic inflation), one attains a plausible picture of a creation mechanism that can underlie an effective multiverse — but at the expense of supposing the validity of untested and perhaps untestable physics. Because of this one does not obtain a specification of a unique multiverse: the physics could be different than what we assumed.

9.2.4 Explanatory power

What explanatory power do we get in return for these problems? It has been suggested they explain the parameters of physics and of cosmology and in particular the very problematic observed value of the cosmological constant [Weinberg, 2000a; Weinberg, 2000b; Susskind, 2005]. The argument goes as follows: assume a multiverse exists; observers can only exist in one of the highly improbable biophilic outliers where the value of the cosmological constant is very small [Hartle, 2004]. A similar argument has been proposed for neutrino masses [Tegmark *et al.*, 2003]. If the multiverse has many varied locations with differing properties, that may indeed help us understand the Anthropic issue: some regions will allow life to exist, others will not [Barrow and Tipler, 1984; Leslie, 1989]. This does provide a useful modicum of explanatory power. However it is far from conclusive. Firstly, it is unclear why the multiverse should have the restricted kinds of variations of the cosmological constant assumed in the various analyses mentioned. If we assume 'all that can happen, happens' the variations will not be of that restricted kind; those analyses will not apply.

Secondly, ultimate issues remain: Why does this unique larger whole have the properties it does? *Why this multiverse rather than any other one*? Why is it a multiverse that allows life to exist? Many multiverses will not allow any life at all. To solve this, we can propose an *ensemble of ensembles of universes*, with even greater explanatory power and even less prospect of observational verification; and so on. The prospect of an infinite regress looms. Indeed if we declare (as suggested at the start of this article) that 'the Universe' is the total of all that physically exists, then when an ensemble of expanding universe domains exists, whether causally connected or not, that ensemble itself should be called 'the Universe', for it is then the totality of physically existing entities. All the foundational problems for a single existing universe domain recur for the multiverse — because when properly considered, it is indeed the Universe!

9.2.5 Testability

If an ensemble exists with members not connected in any physical way to the observable universe, then we cannot interact with them in any way nor observe them, so we can say anything we like about them without fear of disproof.[32] Thus any statements we make about them can have no solid scientific or explanatory status; they are totally vulnerable to anyone else who claims an ensemble with different properties (for example claiming different kinds of underlying logics are possible in their ensemble, or claiming many physically effective gods and devils in many universes in their ensemble).

Thesis H1: Multiverse proposals are unprovable by observation or experiment, but some self-consistency tests are possible. *Direct observations cannot prove or disprove that a multiverse exists, for the necessary causal relations allowing observation or testing of their existence are absent. Their existence cannot be predicted from known physics, because the supposed causal or pre-causal processes are either unproven or indeed untestable. However some self-consistency conditions for specific multiverse models can be tested.*

Any proposed physics underlying a multiverse proposal, such as Coleman-de Luccia tunneling [Coleman and de Luccia, 1980], will be an extrapolation of known physics; but the validity of that major extrapolation to cosmology is untestable.

Attempts have been made to justify the existence of multiverses as testable firstly via Rees' 'slippery slope' argument [Rees, 2003]. This runs as follows: we can reasonably assume galaxies that we cannot see exist outside the visual horizon (Sec. 8.3); why not extend this argument by small steps to show totally disconnected universes exist? The problem is that this assumes a continuity of existence that does not hold good. The domain outside our horizon is assumed to exist with similar properties to those inside because they are a continuous extension of it and have a largely common causal origin; their nature can be inferred from what we can see. Disconnected multiverse domains are assumed to have quite different properties, and their nature cannot be inferred from what we can see as there is no continuity or causal connection.

Secondly, several authors (Leslie [Leslie, 1989], Weinberg [Weinberg, 2000a; Weinberg, 2000b], and Rees [Rees, 2003] for example) have used arguments based on the idea that the universe is no *more* special than it has to be; a form of "speciality argument." According to Rees, if our universe turns out to be *even more specially* tuned than our presence requires, the existence of a multiverse to explain such "over-tuning" would be refuted; but the actual universe is not more special than this, so the multiverse is not refuted.

In more detail: naive quantum physics predicts the cosmological constant Λ to be very large. But our presence in the universe requires it to be small enough that galaxies and stars can form, so Λ must obviously be below that galaxy-forming threshold. If our universe belongs to an ensemble in which Λ was equally likely to take any value in the biophilic region (the uniform probability assumption),[33] then

[32]But there are counter arguments by Leibniz [Wilson, 1989] and Lewis [Lewis, 1986, section 2.4, pp. 108–115].

[33]The probability distribution for Λ will plausibly peak far away from the biophilic region, tailing down to a low value that will be approximately constant in that narrow region, cf.[Hartle, 2004].

we would not expect it to be too far below this threshold. This is because, if it's too far below the threshold, the probability of randomly choosing that universe in the ensemble becomes very small — there are very few universes with such small values of Λ in the biophilic subset of the ensemble. That is, it would be more likely that any bio-friendly universe in the ensemble would have a value of Λ closer to the threshold value. Present data on this value indicates that it is not too far below the threshold. Thus, our universe is not markedly more special that it needs to be as far as Λ is concerned, and so explaining its fine-tuning by existence of a multiverse is legitimate.

Is this argument compelling? It is a reasonable test of consistency for a multiverse that is known to exist, so that probability considerations apply; but they do not apply if there is no multiverse (Sec. 3). Additionally, probability considerations cannot ever be *conclusive*. Indeed,

Thesis H2: Probability-based arguments cannot demonstrate the existence of multiverses. *Probability arguments cannot be used to* prove *the existence of a multiverse, for they are only applicable if a multiverse exists. Furthermore probability arguments can never prove anything for certain, as it is not possible to violate any probability predictions, and this is* a fortiori *so when there is only one case to consider, so that no statistical observations are possible.*

All one can say on the basis of probability arguments is that some specific state is very improbable. But this does not prove it is impossible, indeed if it is stated to have a low probability, that is precisely a statement that it is possible. Thus such arguments can at best only give plausibility indications even when they are applicable. The assumption that probability arguments can be conclusive is equivalent to the claim that the universe is generic rather than special; but whether this is so or not is precisely the issue under debate (see Thesis **D3**). The argument is useful as a plausibility argument for a multiverse, but is not *proof* of its existence.

Finally, it has been proposed that the existence of multiverses is an inevitable consequence of the universe having infinite space sections [Tegmark, 2003; Seife, 2004], because that leads to infinite spatial repetition of conditions (cf. [Ellis and Brundrit, 1979]). But this supposed spatial infinity is an untested philosophical assumption, which certainly cannot be observationally proven to be correct. Apart from the existence of horizons preventing confirmation of this supposition, even if the entire universe were observable, proving it correct would still not be possible because by definition counting an infinite number of objects takes an infinite amount of time. This is an untestable philosophical argument, not an empirically testable one; furthermore, it can be argued to be implausible (Sec. 9.3.2). Indeed current data suggest it is not the case; this is the one good consistency test one can use for some multiverse proposals (Sec. 9.2.7).

9.2.6 Explanation vs Testability

The argument that this infinite ensemble actually exists can be claimed to have a certain explanatory economy, although others would claim that Occam's razor has been completely abandoned in favour of a profligate excess of existential multiplicity, extravagantly hypothesized in order to explain the one universe that we do know exists. Certainly the price is a lack of testability through either observations or experiment — which is usually taken to be an essential element of any serious scientific theory.[34] It is not uniquely definable nor determinable, and there is a complete loss of verifiability. There is no way to determine the properties of any other universe in the multiverse if they do indeed exist, for they are forever outside observational reach. The point is that there is not just an issue of showing a multiverse exists. If this is a scientific proposition one needs to be able to show which specific multiverse exists; but there is no observational way to do this. Indeed if you can't show *which particular* one exists, it is doubtful you have shown *any* one exists.

What does a claim for such existence mean in this context? Gardner puts it this way: "There is not the slightest shred of reliable evidence that there is any universe other than the one we are in. No multiverse theory has so far provided a prediction that can be tested. As far as we can tell, universes are not even as plentiful as even *two* blackberries" [Gardner, 2003].[35]

Thesis H3: Multiverses are a philosophical rather than scientific proposal. *The idea of a multiverse provides a possible route for the explanation of fine tuning. But it is not uniquely defined, is not scientifically testable apart from some possible consistency tests, and in the end simply postpones the ultimate metaphysical questions.*

The definitive consistency tests on some multiverse proposals (Sec. 9.2.7) are *necessary* conditions for those specific multiverse proposals, but are hardly by themselves indications that the multiverse proposal is true. The drive to believe this is the case comes from theoretical and philosophical considerations (see e.g. [Susskind, 2005]) rather than from data. The claim an ensemble physically exists[36] is problematic as a proposal for scientific explanation, if science is taken to involve testability. Indeed, adopting these explanations is a triumph of theory over testability [Gardner, 2003], but the theories being assumed are not testable. It is therefore a metaphysical choice made for philosophical reasons. That does not mean it is unreasonable (it can be supported by quite persuasive plausibility arguments); but its lack of scientific status should be made clear.

[34]In [Stoeger *et al.*, 2004], the framework and conditions under which the multiverse hypothesis would be testable within a retroductive framework, given the rigorous conditions formulated in that paper; are indicated; these conditions are not fulfilled.

[35]This contrasts strongly, for example, with Deutsch's and Lewis's defence of the concept [Deutsch, 1998; Lewis, 1986]. Lewis defends the thesis of "modal realism": that the world we are part of is but one of a plurality of worlds.

[36]As opposed to consideration of an explicitly hypothetical such ensemble, which can indeed be useful, see Sec. 7.2.

9.2.7 Observations and disproof

Despite the gloomy prognosis given above, there are some specific cases where the existence of a chaotic inflation (multi-domain) type scenario (Sec. 2.6) can be *dis*proved. These are firstly when we live in a 'small universe' where we have already seen right round the universe (Sec. 4.3.1), for then the universe closes up on itself in a single FL-like domain, so that no further such domains can exist that are causally connected to us in a single connected spacetime. This 'small universe' situation is observationally testable (Sec. 4.3.1); its confirmation would disprove the usual chaotic inflationary scenario, but not a truly 'disconnected' multiverse proposal, for that cannot be shown to be false by any observation. Neither can it be shown to be true. Secondly, many versions of chaotic inflation, for example those involving Coleman-de Luccia tunneling [Coleman and de Luccia, 1980] from a de Sitter spacetime, demand $k = -1 \Leftrightarrow \Omega_0 < 1$ [Freivogel *et al.*, 2005b; Susskind, 2005]. This requirement is currently marginally disproved by the $2-\sigma$ bounds on Ω_0 when WMAP observations are combined with the best other available data (Sec. 2.3.7). The best current data is marginally consistent with $k = -1$, but the value indicated most strongly by that data is $k = +1$, indicating finite closed space sections rather than an infinite multiverse such as that advocated by Susskind *et al* [Freivogel *et al.*, 2005b; Susskind, 2005].

9.2.8 Physical or biological paradigms — Adaptive Evolution?

Given that the multiverse idea must in the end be justified philosophically rather than by scientific testing, is there a philosophically preferable version of the idea? One can suggest there is: greater explanatory power is potentially available by introducing the major constructive principle of biology into cosmology, namely adaptive evolution, which is the most powerful process known that can produce ordered structure where none pre-existed. This is realized in principle in Lee Smolin's idea (Sec. 9.1.6) of Darwinian adaptation when collapse to black holes is followed by re-expansion, but with an alteration of the constants of physics each time, so as to allow for evolutionary selection towards those regions that produce the maximum number of black holes. The idea needs development, but is very intriguing:

Thesis H4: The underlying physics paradigm of cosmology could be extended to include biological insights. *The dominant paradigm in cosmology is that of theoretical physics. It may be that it will attain deeper explanatory power by embracing biological insights, and specifically that of Darwinian evolution. The Smolin proposal for evolution of populations of expanding universe domains [Smolin, 1992] is an example of this kind of thinking.*

The result is different in important ways from standard cosmological theory precisely because it embodies in one theory three of the major ideas of last century, namely (i) Darwinian evolution of populations through competitive selection, (ii) the evolution of the universe in the sense of major changes in its structure associated with its expansion, and (iii) quantum theory, underlying the only partly explicated mechanism supposed to cause re-expansion out of collapse into a black hole. There is a great contrast with the theoretical physics paradigm of dynamics governed simply by variational principles shaped by symmetry considerations. It seems worth pursuing as a very different route to the understanding of the creation of structure.[37]

9.3 Issue I: Natures of Existence

Underlying all this is the issue of natures of existence, which has a number of aspects, relating from the purely physical to more metaphysical issues.

9.3.1 Physical existence: kinds of matter

Unsolved key issues for physical cosmology relate to what kind of matter and/or fields exist. While we understand matter in the solar system quite well, at present we do not understand most of what exists in the universe at large:

Thesis I1: We do not understand the dominant dynamical matter components of the universe at early or late times. *A key goal for physical cosmology is determining the nature of the inflaton, of dark matter, and of dark energy. Until this is done, the causal understanding of cosmology is incomplete, and in particular the far future fate of the universe is unknown.*

This is the core activity of much work in cosmology at present. Until they are all explicated, cosmology is not properly linked to physics, and the nature of the matter that dominates the dynamics of the universe is unknown. Its explication is surely one of the key concerns of cosmology [Durrer, 2002]. A key requirement is that even if we cannot experimentally verify the proposed nature of the matter, at least it should be physically plausible. This appears not to be the case for some current proposals, e.g. so-called 'phantom matter' which has negative kinetic energy terms.

The far future fate of the universe depends crucially on the effective equation of state for dark matter ('quintessence'). But the problem is that even if we can determine these properties at the present time (for one particular range of parameter values), this does not necessarily guarantee what they will be in the far future (for a quite different range of parameter values that are probably outside the range of possible experimental test). Furthermore adjusting a 'dark energy' model to fit the supernova data does not determine the underlying physics. One can fit any monotonic evolution $S(t)$ with a suitable choice of the equation of state function $p(\mu)$. Specifically, for any $S(t)$ and any k we define $\mu(t)$ and $p(t)$ by

$$(41)\quad \kappa\mu(t) = 3\left[\frac{\dot{S}^2(t)}{S^2(t)} + \frac{k}{S^2(t)}\right],\ \kappa p(t) = \left[\frac{\dot{S}^2(t)}{S^2(t)} + \frac{k}{S^2(t)}\right] - 2\frac{\ddot{S}(t)}{S(t)},$$

[37] Cf. Chapter 13 of Susskind [Susskind, 2005].

then (9), (7) will be exactly satisfied, and we have 'solved' the field equations for this arbitrarily chosen monotonic evolution $S(t)$. If we can observationally determine the form of $S(t)$, for example from (m, z)–curves associated with supernovae data, this is essentially how we can then determine that some kind of 'dark energy' or 'quintessence' is required to give that evolution, and we can find the equation of state implied by eliminating t between these two equations. This is, however, not a *physical* explanation until we have either in some independent experimental test demonstrated that matter of this form exists, or have theoretically shown why this matter or field has the form it does in some more fundamental terms than simply a phenomenological fit. If we assume the matter is a scalar field, the kinetic energy term $\dot{\phi}^2$ implied by (32), (41) may be negative — which is the case for so-called 'shadow matter' models proposed recently by some worker. If normal physics criteria are applied, this is a proof that this kind of matter is unphysical, rather than an identification of the nature of the dark energy.

9.3.2 Existence of Infinities

The nature of existence is significantly different if there is a finite amount of matter or objects in the universe, as opposed to there being an infinite quantity in existence. Some proposals claim there may be an infinite number of universes in a multiverse and many cosmological models have spatial sections that are infinite, implying an infinite number of particles, stars, and galaxies. However, infinity is quite different from a very large number! Following David Hilbert [Hilbert, 1964], one can suggest these unverifiable proposals cannot be true: the word 'infinity' denotes a quantity or number that can never be attained, and so will never occur in physical reality.[38] He states

"Our principal result is that the infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought . . . The role that remains for the infinite to play is solely that of an idea . . . which transcends all experience and which completes the concrete as a totality . . ." [Hilbert, 1964, p. 151].

This suggests "infinity" cannot be arrived at, or realized, in a concrete physical setting; on the contrary, the concept itself implies its inability to be realized![39]

Thesis I2: The often claimed physical existence of infinities is questionable. *The claimed existence of physically realized infinities in cosmology or multiverses raises problematic issues. One can suggest they are unphysical; in any case such claims are certainly unverifiable.*

This applies in principle to both small and large scales in any single universe:

[38] An intriguing further issue is the dual question: Does the quantity zero occur in physical reality? This is related to the idea of physical existence of nothingness, as contrasted with a vacuum [Seife, 2000]. A vacuum is not nothing! (cf. [Susskind, 2005]).

[39] For a contrasting view, see Bernadete [Bernadete, 1964].

- The existence of a physically existing spacetime continuum represented by a real (number) manifold at the micro-level contrasts with quantum gravity claims of a discrete spacetime structure at the Planck scale, which one might suppose was a generic aspect of fully non-linear quantum gravity theories [Rovelli, 2004]. In terms of physical reality, this promises to get rid of the uncountable infinities the real line continuum engenders in all physical variables and fields.[40] There is no experiment that can *prove* there is a physical continuum in time or space; all we can do is test space-time structure on smaller and smaller scales, but we cannot approach the Planck scale.

- Infinitely large space-sections at the macro-level raise problems as indicated by Hilbert, and leads to the infinite duplication of life and all events [Ellis and Brundrit, 1979]. We may assume space extends forever in Euclidean geometry and in many cosmological models, but we can never prove that any realised 3-space in the real universe continues in this way — it is an untestable concept, and the real spatial geometry of the universe is almost certainly not Euclidean. Thus Euclidean space is an abstraction that is probably not physically real. The infinities supposed in chaotic inflationary models derive from the presumption of pre-existing infinite Euclidean space sections, and there is no reason why those should necessarily exist. In the physical universe spatial infinities can be avoided by compact spatial sections, resulting either from positive spatial curvature, or from a choice of compact topologies in universes that have zero or negative spatial curvature. Machian considerations to do with the boundary conditions for physics suggest this is highly preferable [Wheeler, 1968]; and if one invokes string theory as a fundamental basis for physics, then 'dimensional democracy' suggests the three large spatial dimensions should also be compact, since the small ('compactified') dimensions are all taken to be so. The best current data from CBR and other observations (Sec. 2.3.7) indeed suggest $k = +1$, implying closed space sections for the best-fit FL model.

- The existence of an eternal universe implies that an infinite time actually exists, which has its own problems: if an event happens at any time t_0, one needs an explanation as to why it did not occur before that time (as there was an infinite previous time available for it to occur); and Poincaré eternal return (mentioned in Sec. 6.1) will be possible if the universe is truly cyclic. In any case it is not possible to *prove* that the universe as a whole, or even the part of the universe in which we live, is past infinite; observations cannot do so, and the physics required to guarantee this would happen (if initial conditions were right) is untestable. Even attempting to prove it is future infinite is problematic (we cannot for example guarantee the properties of the

[40] To avoid infinities entirely would require that nothing whatever is a continuum in physical reality (since any continuum interval contains an infinite number of points). Doing without that, conceptually, would mean a complete rewrite of many things. Considering how to do so in a way compatible with observation is in my view a worthwhile project.

vacuum into the infinite future — it might decay into a state corresponding to a negative effective cosmological constant).

- It applies to the possible nature of a multiverse. Specifying the geometry of a generic universe requires an infinite amount of information because the quantities necessary to do so are fields on spacetime, in general requiring specification at each point (or equivalently, an infinite number of Fourier coefficients): they will almost always not be algorithmically compressible. All possible values of all these components in all possible combinations will have to occur in a multiverse in which "all that can happen, does happen". There are also an infinite number of topological possibilities. This greatly aggravates all the problems regarding infinity and the ensemble. Only in highly symmetric cases, like the FL solutions, does this data reduce to a finite number of parameters, each of which would have to occur in all possible values (which themselves are usually taken to span an infinite set, namely the entire real line). Many universes in the ensemble may themselves have infinite spatial extent and contain an infinite amount of matter, with all the problems that entails. To conceive of physical creation of an infinite set of universes (most requiring an infinite amount of information for their prescription, and many of which will themselves be spatially infinite) is at least an order of magnitude more difficult than specifying an existent infinitude of finitely specifiable objects.

One should note here particularly that problems arise in the multiverse context from the continuum of values assigned by classical theories to physical quantities. Suppose for example that we identify corresponding times in the models in an ensemble and then assume that *all* values of the density parameter and the cosmological constant occur at each spatial point at that time. Because these values lie in the real number continuum, this is a doubly uncountably infinite set of models. Assuming genuine physical existence of such an uncountable infinitude of universes is the antithesis of Occam's razor. But on the other hand, if the set of realised models is either finite or countably infinite, then almost all possible models are not realised. And in any case this assumption is absurdly unprovable. We can't observationally demonstrate a single other universe exists [Gardner, 2003], let alone an infinitude. The concept of infinity is used with gay abandon in some multiverse discussions [Knobe *et al.*, 2005], without any concern either for the philosophical problems associated with this statement [Hilbert, 1964], or for its completely unverifiable character. It is an extravagant claim that should be treated with extreme caution.

9.3.3 The Nature of the Laws of Physics

Underlying all the above discussion is the basic concept of ordered behaviour of matter, characterized by laws of physics of a mathematical nature that are the same everywhere in the universe.[41] Three interlinked issues arise.

(i) *What is the ontological nature of the laws of physics*: descriptive, *just characterizing the way things are, or* prescriptive, *enforcing them to be this way*? [Carroll, 2004]. If they are descriptive, the issue arising is, *Why does all matter have the same properties wherever it exists in the universe*? Why are all electrons everywhere in the universe identical, if the laws are only descriptive? If they are prescriptive, then matter will necessarily be the same everywhere (assuming the laws themselves are invariable); the issue arising then is, *In what way do laws of physics exist that enforce themselves on the matter in the universe*? Do they for example have an existence in some kind of Platonic space that controls the nature of matter and existence? One can avoid talking about the laws of physics *per se* by instead considering the *space of possibilities* underlying what exists physically, rigorously constraining the possible natures of what actually comes into existence [Ellis, 2004]. This space is more or less uniquely related to the underlying laws in the same way that the space of solutions of differential equations is related to the nature of the equations. This enables one to avoid the issue of the ontology of the laws of physics, but does not solve it.

(ii) *Why are the laws of physics so well explained by mathematical descriptions*? If they are prescriptive, this deep issue might be related to the suggested Platonic nature of the space of mathematical reality [Penrose, 2004]. If they are descriptive, then the mathematical expressions we use to encapsulate them are just a convenient description but do not reflect their ultimate nature. Many writings in physics and cosmology seem to assume that their ultimate existential nature is indeed mathematical — perhaps a confusion of appearance and reality (see Sec. 8.3).

(iii) *Do they pre-exist the universe and control its coming into being, or do they come into being with the universe*? This is where this issue relates deeply to the nature of cosmology, and is clearly related to the other two questions raised above. Many theories of creation of the universe assume that all these laws, or at least a basic subset, pre-exist the coming into being of the physical universe, because they are presumed to underlie the creation process, for example the entire apparatus of quantum field theory is often taken for granted as pre-existing our universe (Sec. 6). This is of course an unprovable proposition

Thesis I3: A deep issue underlying the nature of cosmology is the nature of the laws of physics. *The nature of the possibility space for physical existence is characterized by the laws of physics. However it is unclear if these laws are prescriptive or descriptive; whether they come into being with space-time and matter, or pre-exist them.*

[41]The effective laws may vary from place to place because for example the vacuum state varies [Susskind, 2005]; but the fundamental laws that underlie this behaviour are themselves taken to be invariant.

9.3.4 'Ultimate Reality'

Philosophers have debated for millennia whether the ultimate nature of existence is purely material, or embodies some form of rationality ('Logos') and/or purpose ('Telos'). *What in the end underlies it all*? Is the ultimate nature of the universe purely material, or does it in some way have an element of the mental? (cf. Sec. 9.1.6). That profound debate is informed by physical cosmology, but cannot be resolved by the physical sciences alone (Sec. 9.1.7). Here, I will make just two comments on this deep issue.

Firstly, even in order to understand just the material world, it can be claimed that one needs to consider forms of existence other than the material only — for example a Platonic world of mathematics and a mental world, both of which can be claimed to exist and be causally effective in terms of affecting the material world [Ellis, 2004; Penrose, 2004]. Our understanding of local causation will be incomplete unless we take them into account.

Secondly, in examining these issues one needs to take into account data about the natures of our existence that come from our daily lives and the broad historical experience of humanity (our experiences of ethics and aesthetics, for example), as well as those discoveries attained by the scientific method. Many writings claim there is no purpose in the universe: it is all just a conglomerate of particles proceeding at a fundamental level in a purposeless and meaningless algorithmic way. But I would reply, the very fact that those writers engage in such discourse undermines their own contention; they ignore the evidence provided by their own actions. There is certainly meaning in the universe to this degree: *the fact they take the trouble to write such contentions is proof that they consider it meaningful to argue about such issues*; and this quality of existence has emerged out of the nature of the physical universe (Sec. 7.3). Indeed the human mind is causally effective in the real physical world precisely through many activities motivated by meanings perceived by the human mind. Any attempt to relate physics and cosmology to ultimate issues must take such real world experience seriously [Ellis, 2005], otherwise it will simply be ignoring a large body of undeniable data. This data does not resolve the ultimate issues, but does indicate dimensions of existence that indeed do occur.

10 CONCLUSION

The physical scale of the Universe is enormous, and the images of distant objects from which we obtain our information are extremely faint. It is remarkable that we are able to understand the Universe as well as we do. An intriguing feature is the way in which the philosophy of cosmology is to a considerable degree shaped by contingent aspects of the nature of the universe — its vast scale (Sec. 4), leading to the existence of visual horizons (Sec. 4.3), and the occurrence of extreme energies in the early universe (Sec. 5), leading to the existence of physical horizons. Philosophical issues arising in relation to cosmology (Sec. 8) would be quite different if its physical structure were very different. Furthermore in order that philosophical analysis can engage with cosmology in depth, the detailed nature of the relation between observations and theory in cosmology (Sec. 2) is relevant.

10.1 Are there laws of cosmology?

As we have discussed in detail, the uniqueness of the universe implies the unique nature of cosmology. We now return to the initial issue, *Are there Laws of the Universe*? (Sec. 3). At one level, the laws of the cosmos are simply the local laws we know and love (e.g. Maxwell's laws, Einstein's field equations) applied to the whole shebang. Of course, there is the problem of extrapolation from the local to the global. But although the extrapolation is bigger in cosmology, it seems not to be different in kind from what we always do in science. In that sense, there are no special laws for the evolution of the universe. But that does not determine the outcome: cosmology needs some prescription of boundary or initial conditions as well, in order to determine the future. Is there a true "Cosmological principle", a law of initial conditions for the universe, that determines this outcome?

The idea of "Laws of initial conditions of the universe" seems not to be a testable idea (Sec. 3). Scientifically, one can only describe what occurred rather than relate it to generic principles, for such principles cannot be tested. In fact any description of boundary or initial conditions for the universe seems to be just that: a description of these conditions, rather than a testable prescription of how they must be. The 'Cosmological Principle' — the universe is necessarily spatially homogeneous and isotropic (Sec. 4.2.2) is of this kind: a description of the way the initial data turned out, rather than a fundamental reason for why this should be so. Justification of this view was based by some workers on a *Copernican Principle* (the assumption we do not live in a privileged place in the universe), strengthened to become a *Cosmological Principle* [Bondi, 1960; Weinberg, 1972; Harrison, 2000]; but this is a philosophical assumption — essentially, a claim that the universe must have very special initial conditions — which may or may not be true, and does not attempt a physical explanation. This kind of argument is out of fashion at present, because we now prefer generality to speciality and physical argumentation to geometrical prescription; but it was previously strongly proposed (e.g. [Weinberg, 1972], pp. 407-412). The tenor of philosophical argument has changed.

Nevertheless there is one kind of Law of the Universe one might propose, following McCrea [McCrea, 1970]: namely an "Uncertainty principle in cosmology", dual to the uncertainty principle in quantum theory. Uncertainty applies on the largest scale, as we have discussed above in some detail, and also on the smallest, where it is a profound feature of quantum theory. Its basis is very different in the two cases, on the one hand (in quantum theory) being ontological in nature, on the other (in cosmology) being epistemological in nature.[42] Nevertheless it is

[42]Assuming that quantum uncertainty is indeed ontological rather than epistemological. One should however keep an open mind on this: just because it is the current dogma does not

a key aspect of our relation to the cosmos, so that (following McCrea) we might perhaps formalize it in order to emphasize its centrality to the relation between cosmology and philosophy:

Thesis of Uncertainty: Ultimate uncertainty is a key aspect of cosmology. *Scientific exploration can tell us much about the universe but not about its ultimate nature, or even much about some of its major geometrical and physical characteristics. Some of this uncertainty may be resolved, but much will remain. Cosmological theory should acknowledge this uncertainty.*

10.2 What can we truly claim

Cosmology considers questions of physical origins in the uniquely existing physical universe (Sec. 6) which provides the context of our existence (Sec. 7, Sec. 9.1). These questions can be extended to include ultimate issues if we so desire (Sec. 8.2), but physical theory cannot resolve them (Sec. 9.1.7). In the end, there are a variety of mysteries underlying the existence and nature of the universe (Sec. 9.3). The scientific study of cosmology can help illuminate their nature, but cannot resolve them.

As well as celebrating the achievements of cosmology, one should fully take into account the limits and problems considered in this chapter, and not claim for scientific cosmology more than it can actually achieve or more certainty than is in fact attainable. Such claims will in the long term undermine cosmology's legitimate status as a project with solid scientific achievements to its name. That status can be vigorously defended as regards the 'Standard Model' of cosmology (Sec. 2.8), provided this standard model is characterized in conservative terms so that it is not threatened by relatively detailed shifts in theory or data that do not in fact threaten the core business of cosmology. Further, this defence must take adequate cognisance of the difficult philosophical issues that arise if one pushes the explanatory role of cosmological theory to its limits (Sec. 6); for example one should not make too strong *scientific* claims in regard to the possible existence of multiverses (Sec. 9.2); philosophically based plausibility arguments for them are fine, if identified as such. Cosmology is not well served by claims that it can achieve more explanatory power than is in fact attainable, or by statements that its claims are verified when in fact the requisite evidence is unavailable, and in some cases must forever remain so.

ACKNOWLEDGEMENT

I thank Bill Stoeger, Martin Bojowald, Malcolm MacCallum, Henk van Elst, Jeremy Butterfield, and John Earman for useful comments that have improved this article.

---

necessarily mean it is true.

Abbreviations used:

CBR: Cosmic Blackbody Radiation, p. 1192
CDM: Cold Dark Matter, p. 1201
EFE: Einstein Field Equations, p. 1185
FL: Friedmann-Lemaître (universe models), p. 1187
HBB: Hot Big Bang, p. 1191
LSS: Last Scattering surface, p. 1192
RW: Robertson-Walker (geometry), p. 1186
SAP: Strong Anthropic Principle, p. 1254
WAP: Weak Anthropic Principle, p. 1253

BIBLIOGRAPHY

[Allday, 2002] J. Allday. *Quarks, Leptons and the Big Bang.* Institute of Physics, 2002.

[Balashov, 1991] Y. V. Balashov. Resource letter Ap-1: The anthropic principle. *American Journal of Physics*, 54:1069–1076, 1991.

[Barbour and Pfister, 1995] J. Barbour and H. Pfister. *Mach's Principle: From Newton's Bucket to Quantum Gravity.* Birkhäuser, Basel, 1995.

[Bardeen, 1980] J. M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev.*, D22:1882-1905, 1980.

[Barnett *et al.*, 1996] R. M. Barnett *et al.*(Particle Data Group) Particle physics summary. a digest of the 1996 Review of Particle Physics. *Reviews of Modern Physics*, 68:611–732, 1996.

[Barrow and Tipler, 1984] J. Barrow and F. Tipler. *The Cosmological Anthropic Principle.* Oxford University Press, 1984.

[Barrow, 2003] J. D. Barrow. *The Constants of Nature.* Vintage, 2003.

[Bennet et al., 2003] C. L. Bennet *et al.* (the WMAP team). First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. *Astrophysical Journal Supplement Series*, 148:1–27, 2003. astro-ph/0302207.

[Berendzen *et al.*, 1976] R. Berendzen, R. Hart and D. Seeley. *Man Discovers the Galaxies.* Science History Publications, New York, 1976.

[Bernadete, 1964] J. A. Bernadete. *Infinity, An Essay in Metaphysics.* Clarendon Press, Oxford, 1964.

[Blau and Guth, 1987] S. K. Blau and A. H. Guth. Inflationary cosmology. In Hawking, S. W. and Israel, W., editors, *300 Years of Gravitation*, page 524. Cambridge University Press, 1987.

[Bojowald, 2001] M. Bojowald. Absence of singularity in loop quantum cosmology. *Physical Review Letters*, 86:5227–5239, 2001.

[Bondi, 1947] H. Bondi. Spherically symmetric models in general relativity. *Monthly Notice of the Royal Astronomical Society*, 107:410, 1947.

[Bondi, 1960] H. Bondi. *Cosmology.* Cambridge University Press, Cambridge, 1960.

[Bonnor and Ellis, 1986] W. B. Bonnor and G. F. R. Ellis. Observational homogeneity of the universe. *Monthly Notice of the Royal Astronomical Society*, 218:605–614, 1986.

[Bostrom, 2002] N. Bostrom. *Anthropic bias: Observation selection effects in science and philosophy.* Routledge, New York, 2002.

[Bothun, 1998] G. Bothun. *Modern Cosmological Observations and Problems.* Taylor and Francis, 1998.

[Brading and Castellani, 2006] K. Brading and E. Castellani. This vol, Ch. 13, 2006.

[Brans and Dicke, 1961] C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Physical Review*, 124:925, 1961.

[Carr and Rees, 1979] B. J. Carr and M. J. Rees. The anthropic principle and the structure of the physical world. *Nature*, 278:605–612, 1979.

[Carroll, 2004] J. Carroll, editor. *Readings on Laws of Nature.* University of Pittsburgh Press, 2004.

[Carroll and Chen, 2005] S. M. Carroll and J. Chen. Does inflation provide natural initial conditions for the universe? gr-qc/0505037, 2005.

[Challinor and Lasenby, 1998] A. D. Challinor and A. N. Lasenby. Cosmic microwave background anisotropies in the CDM model: A covariant and gauge-invariant approach. *Physical Review D*, 58. astro-ph/9804301, 1998.

[Coleman and de Luccia, 1980] S. Coleman and F. de Luccia. Gravitational effects on and of vacuum decay. *Physical Review D*, 21:3305–3315, 1980.

[Copeland *et al.*, 2005] E. J. Copeland, J. E. Lidsey, and S. Mizuno. Correspondence between loop-inspired and braneworld cosmology. gr-qc/0510022, 2005.

[Cowie and Songaila, 1995] L. Cowie and A. Songaila. Astrophysical limits on the evolution of dimensionless physical constants over cosmological time. *Astrophysics Journal*, 453:596–598, 1995.

[Davies, 1974] P. C. W. Davies. *The physics of time asymmetry.* Surrey University Press, London, 1974.

[Davies, 1982] P. C. W. Davies. *The Accidental Universe.* Cambridge University Press, 1982.

[Davies, 1987] P. C. W. Davies. *The cosmic Blueprint.* Heinemann, London, 1987.

[Davis and Lineweaver, 2004] T. M. Davis. and C. H. Lineweaver. Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Publications of the Astronomical Society of Australia*, 21:97–109, 2004. astro-ph/0310808.

[de Vaucouleurs, 1970] G. de Vaucouleurs. The case for a hierarchical cosmology. *Science*, 167:1203–13, 1970.

[Deutsch, 1998] D. Deutsch. *The Fabric of reality: The science of parallel universes - and its implications.* Penguin, 1998.

[Dicke and Peebles, 1979] R. H. Dicke. and P. J. E. Peebles. The big bang cosmology—enigmas and nostrums. In Hawking, S. W. and Israel, W., editors, *General Relativity: An Einstein Centenary Survey*, pages 504–517. Cambridge University Press, Cambridge, 1979.

[Dickson, 2006] M. I. Dickson. this vol , Ch.4, 2006.

[Dirac, 1938] P. A. M. Dirac. New basis for cosmology. *Proceedings of the Royal Society of LondoAn*, 165:199–208, 1938.

[Disney, 1976] M. J. Disney. Visibility of galaxies. *Nature*, 263:573–575, 1976.

[Dodelson, 2003] S. Dodelson. *Modern Cosmology.* Academic Press, 2003.

[Durrer, 2002] R. Durrer. Frontiers of the universe: What do we know and what do we understand? astro-ph/0205101, 2002.

[Earman, 1987] J. Earman. The SAP also rises: A critical examination of the anthropic principle. *American Philosophical Quarterly*, 24:307–325, 1987.

[Earman, 1999] J. Earman. The Penrose-Hawking singularity theorems: History and implications. In Goenner, H., Renn, J., and Sauer, T., editors, *The Expanding Worlds of General Relativity.* Birkhäuser, 1999.

[Earman, 2001] J. Earman. Lambda: The constant that would not die. *Archive for History of Exact Sciences*, 55:189–220, 2001.

[Earman, 2003] J. Earman. The cosmological constant, the fate of the universe, unimodular gravity, and all that. *Studies in History and Philosophy of Modern Physics*, 34:559–577, 2003.

[Earman and Mosterin, 1999] J. Earman and J. Mosterin. A critical look at inflationary cosmology. *Philosophy of Science*, 66:1–49, 1999.

[Ehlers, 1993] J. Ehlers. Contributions to the relativistic mechanics of continuous media. *General Relativity and Gravitation*, 25:1225–1266, 1993.

[Ehlers *et al.*, 1968] J. Ehlers, P. Geren, and R. K. Sachs. Isotropic solutions of the Einstein-Liouville equations. *Journal of Mathematical Physics*, 9:1344–1349, 1968.

[Ehlers and Rindler, 1989] J. Ehlers and W. Rindler. A phase-space representation of Friedmann-Lemaître universes containing both dust and radiation and the inevitability of a big bang. *Monthly Notice of the Royal Astronomical Society*, 238:503–521, 1989.

[Eisenstein *et al.*, 2005a] D. J. Eisenstein *et al.* Detection of the baryon accoustic peak in the large scale correlation function of galaxies, 2005. astro-ph/0501171.

[Ellis, 1967] G. F. R. Ellis. The dynamics of pressure-free matter in general relativity. *Journal of Mathematical Physics*, 8:1171–1194, 1967.

[Ellis, 1971a] G. F. R. Ellis. Relativistic cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School.* Academic Press, New York, 1971.

[Ellis, 1971b] G. F. R. Ellis. Topology and cosmology. *General Relativity and Gravitation*, 2:7–21, 1971.

[Ellis, 1975] G. F. R. Ellis. Cosmology and verifiability. *Quarterly Journal of the Royal Astronomical Society*, 16:245–264, 1975.

[Ellis, 1980] G. F. R. Ellis. Limits to verification in cosmology. *Ann New York Acad Sci*, 336:130–160, 1980.

[Ellis, 1984] G. F. R. Ellis. Relativistic cosmology: its nature, aims and problems. In *et al.*, B. B., editor, *General Relativity and Gravitation*, pages 215–288. Reidel, 1984.

[Ellis, 1989] G. F. R. Ellis. A history of cosmology 1917-1955. In Howard, D. and Stachel, J., editors, *Einstein and the History of General Relativity*, volume 1 of *Einstein Study Series*, pages 367–431. Birkhäuser, Boston, 1989.

[Ellis, 1990] G. F. R. Ellis. Innovation resistance and change: the transition to the expanding universe. In Bertotti, B., Balbinto, R., Bergia, S., and Messina., A., editors, *Modern Cosmology in Retrospect*, pages 97–114. Cambridge University Press, 1990.

[Ellis, 1991] G. F. R. Ellis. Major themes in the relation between philosophy and cosmology. *Mem Ital Ast Soc*, 62:553–605, 1991.

[Ellis, 2002] G. F. R. Ellis. Cosmology and local physics. *New Astronomy Reviews*, 46:645–658, 2002. gr-qc/0102017.

[Ellis, 2004] G. F. R. Ellis. True complexity and its associated ontology. In Barrow, J. D., Davies, P. C., , and Jr., C. L. H., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity, honoring John Wheeler's 90th birthday.* Cambridge University Press, 2004.

[Ellis, 2005] G. F. R. Ellis. Physics, complexity, and causality. *Nature*, 435:743, 2005.

[Ellis and Baldwin, 1984] G. F. R. Ellis. and J. Baldwin. On the expected anisotropy of radio source counts. *Monthly Notice of the Royal Astronomical Society*, 206:377–381, 1984.

[Ellis and Brundrit, 1979] G. F. R. Ellis and G. B. Brundrit. Life in the infinite universe. *Quarterly Journal of the Royal Astronomical Society*, 20:37–41, 1979.

[Ellis and Bruni, 1989] G. F. R. Ellis and M. Bruni. A covariant and gauge-free approach to density fluctuations in cosmology. *Physical Review D*, 40:1804–1818, 1989.

[Ellis and Buchert, 2006] G. F. R. Ellis and T. R. Buchert. The universe seen at different scales. To appear, *Physics Letter A*, special Einstein issue, 2006. gr-qc/0506106.

[Ellis and King, 1974] G. F. R. Ellis and A. R. King. Was the big bang a whimper? *Communications in Mathematical Physics*, 38:119–165, 1974.

[Ellis *et al.*, 2004] G. F. R. Ellis, U. Kirchner, and W. Stoeger. Multiverses and physical cosmology. *Monthly Notice of the Royal Astronomical Society*, 347:921–936, 2004.

[Ellis and Maartens, 2004] G. F. R. Ellis and R. Maartens. The emergent universe: inflationary cosmology with no singularity and no quantum gravity era. *Classical and Quantum Gravity*, 21:223–232, 2004.

[Ellis *et al.*, 1978] G. F. R. Ellis, R. Maartens, and S. D. Nel. The expansion ofthe universe. *Monthly Notice of the Royal Astronomical Society*, 184:439–465, 1978.

[Ellis and Madsen, 1991] G. F. R. Ellis and M. Madsen. Exact scalar field cosmologies. *Classical and Quantum Gravity*, 8:667–676, 1991.

[Ellis *et al.*, 2002a] G. F. R. Ellis, P. McEwan, W. Stoeger, and P. Dunsby. Causality in inflationary universes with positive spatial curvature. *General Relativity and Gravitation.*, 34:1461–1481, 2002. gr-qc/0109024.

[Ellis *et al.*, 1985] G. F. R. Ellis, S. D. Nel, W. Stoeger, R. Maartens, and A. P. Whitman. Ideal observational cosmology. *Physics Reports*, 124:315–417, 1985.

[Ellis *et al.*, 1984] G. F. R. Ellis, J. J. Perry, and A. Sievers. Cosmological observations of galaxies: the observational map. *Astronomical Journal*, 89:1124–1154, 1984.

[Ellis and Schreiber, 1986] G. F. R. Ellis and G. Schreiber. Observational and dynamic properties of small universes. *Physics Letters A*, 115:97–107, 1986.

[Ellis and Sciama, 1972] G. F. R. Ellis and D. W. Sciama. Global and non-global problems in cosmology. In O'Raifeartaigh, L., editor, *General Relativity*, pages 35–59. Oxford University Press, 1972.

[Ellis *et al.*, 2002b] G. F. R. Ellis, W. Stoeger, P. McEwan, and P. Dunsby. Dynamics of inflationary universes with positive spatial curvature. *General Relativity and Gravitation*, 34:1445–1459, 2002. gr-qc/0109023.

[Ellis and Stoeger, 1987] G. F. R. Ellis and W. R. Stoeger. The Fitting problem in cosmology. *Classical and Quantum Gravity*, 4:1679–1690, 1987.

[Ellis and Stoeger, 1988] G. F. R. Ellis and W. R. Stoeger. Horizons in inflationary universes. *Classical and Quantum Gravity*, 5:207, 1988.

[Ellis and Uzan, 2005] G. F. R. Ellis and J.-P. Uzan. 'c' is the speed of light, isn't it? *American Journal of Physics*, 73:240–247, 2005.

[Ellis and van Elst, 1999a] G. F. R. Ellis and H. van Elst. Cosmological models (cargese lectures 1998). In Lachièze-Ray, M., editor, *Theoretical and Observational Cosmology*, volume 541, pages 1–116. Kluwer, Nato Series C: Mathematical and Physical Sciences, 1999. gr-qc/9812046.

[Ellis and van Elst, 1999b] G. F. R. Ellis and H. van Elst. Deviation of geodesics in FLRW spacetime geometries. In Harvey, A., editor, *On Einstein's Path: Essays in Honour of Englebert Schucking*, pages 203–226. Springer, 1999.

[Ellis and Williams, 2000] G. F. R. Ellis and R. M. Williams. *Flat and Curved Spacetimes*. Cambridge Univesity Press, 2000.

[Fabian, 1989] A. C. Fabian, editor. *Origins*. Cambridge University Press, 1989.

[Freivogel *et al.*, 2005a] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Freivogel *et al.*, 2005b] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Gardner, 1986] M. Gardner. WAP, SAP, PAP, and FAP. *The New York Review of Books*, 23(8):22–25, 1986.

[Gardner, 2003] M. Gardner. *Are Universes Thicker than Blackberries?* W W Norton, 2003.

[Garrett and Coles, 1993] A. J. M. Garrett and P. Coles. Bayesian inductive inference and the anthropic cosmological principle. *Comments Astrophys*, 17:23–47, 1993.

[Ghosh *et al.*, 1998] A. Ghosh, G. Pollifrone, and G. Veneziano. Quantum fluctuations in open pre-big bang cosmology. *Physics LetterBs*, 440:20–27, 1998.

[Gibbons *et al.*, 1983] G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos. *The Very Early Universe*. Cambridge University Press, 1983.

[Gibbons *et al.*, 2003] G. W. Gibbons, E. P. S. Shellard, and S. J. Rankin, editors. *The Future of Theoretical Physics and Cosmology: Celebrating Stephen Hawking's 60th Birthday*. Cambridge University Press, 2003.

[Goodman, 1995] J. Goodman. Geocentrism reexamined. *Physical Review D*, 52:1821–1827, 1995.

[Gott and Li, 1998] J. R. Gott and L.-X. Li. Can the universe create itself? *Physical Review D*, 58:023501, 1998. astro-ph/9712344.

[Gribbin and Rees, 1991] J. Gribbin and M. Rees. *Cosmic Coincidences*. Black Swan, 1991.

[Grunbaum, 1989] A. Grunbaum. The pseudo problem of creation. *Philosophy of Science*, 56:373–394, 1989. reprinted in [Leslie, 1998].

[Grunbaum, 2004] A. Grunbaum. The poverty of theistic cosmology. *British Journal of Philosophy of Science*, 55:561–614, 2004.

[Gunn *et al.*, 1978] J. E. Gunn, M. S. Longair and M. J. Rees. *Observational cosmology: Saas-Fee Course 8*. Geneva Observatory, Switzerland, 1978.

[Guth, 1997] A. H. Guth. *The Inflationary Universe: The Quest for a new Theory of Cosmic Origins*. Addison Wesley, 1997.

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

[Guth, 2001] A. H. Guth. Eternal inflation. In *Proceedings of "Cosmic Questions"Meeting*, volume 95. The New York Academy of Sciences Press, 2001. astro-ph/0101507.

[Harris *et al.*, 1998] W. E. Harris, P. R. Durell, M. J. Pierce, and J. Seckers. Constraints in the Hubble constant from observations of the brightest red giant stars in a Virgo-cluster galaxy. *Nature*, 395:45, 1998.

[Harrison, 2000] E. R. Harrison. *Cosmology: The Science of the Universe*. Cambridge University Press, Cambridge, 2 edition, 2000.

[Hartle, 2004] J. B. Hartle. Anthropic reasoning and quantum cosmology. In Allen, R. and Pope, C., editors, *The New Cosmology*. AIP, 2004. gr-qc/0406104.

[Harvey, 2006] J. Harvey. this vol, Ch. 14, 2006.

[Harwit, 1984] M. Harwit. *Cosmic Discovery*. MIT Press, 1984.

[Hawking, 1987] S. W. Hawking. Quantum cosmology. In Hawking, S. W. and Israel, W., editors, *300 Years of Gravitation*, page 631. Cambridge University Press, Cambridge, 1987.

[Hawking, 1993] S. W. Hawking. *On the Big Bang and Black Holes*. World Scientific, 1993.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-time*. Cambridge University Press, 1973.

[Hilbert, 1964] D. Hilbert. On the infinite. In Benacerraf, P. and Putnam, H., editors, *Philosophy of Mathematics*, pages 134–151. Englewood Cliff, N J.: Prentice Hall, 1964.
[Hogan, 1999] C. J. Hogan. Where the baryons are. astro-ph/9912107, 1999.
[Hogan, 2003] C. J. Hogan. Why the universe is just so. *Reviews of Modern Physics*, 72:1149–1161, 2003.
[Hollands and Wald, 2002] S. Hollands and R. M. Wald. An alternative to inflation. *Gen Rel.Grav*, 34:2043–2055, 2002.
[Hoyle, 1960] F. Hoyle. Cosmological tests of gravitational theories. In *Rendiconti Scuola Enrico Fermi XX Corso*, pages 141–. Academic Press, New York, 1960.
[Hu and Sugiyama, 1995a] W. Hu and N. Sugiyama. Anisotropies in the CMB: An analytic approach. *Astrophysical Journal*, 444:489–506, 1995.
[Hu and Sugiyama, 1995b] W. Hu and N. Sugiyama. Toward understanding CMB anisotropies and their implications. *Physical Review D*, 51:2599–2630, 1995.
[Hubble, 1936] E. Hubble. *The Realm of the Nebulae*. Yale University Press, 1936.
[Hume, 1993] D. Hume. *Dialogues and Natural History of Religion*. World Classics. Oxford University Press, 1993.
[Isham, 1997] C. J. Isham. *Lectures on Quantum Theory: Mathematical and Structural Foundations*. Imperial College Press, London, 1997.
[Jain and Dev, 2005] D. Jain and A. Dev. Age of high redshift objects - a litmus test for dark energy models. astro-ph/0509212, 2005.
[Kamionkowski and Loeb, 1997] M. Kamionkowski and A. Loeb. Getting around cosmic variance. *Physical Review D*, 56:4511, 1997.
[Kantowski, 1998] R. Kantowski. The effects of inhomogeneities on evaluating the mass parameter ω_m and the cosmological constant λ. *Astrophysical Journal*, 507:483–496, 1998.
[Kantowski *et al.*, 1995] R. Kantowski, T. Vaughan, and D. Branch. The effects of inhomogeneities on evaluating the deceleration parameter. *Astrophysical Journal*, 447:35–42, 1995.
[Khoury *et al.*, 2001] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok. The ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Physical Review D*, 64:123522, 2001.
[Knobe *et al.*, 2005] J. Knobe, K. D. Olum, and A. Vilenkin. Philsophical implications of inflationary cosmology. phys/0320071, 2005.
[Kolb and Turner, 1990] E. W. Kolb and M. S. Turner. *The Early Universe*. Addison Wesley, 1990.
[Kragh, 1996] H. Kragh. *Cosmology and Controversy*. Princeton University Press, Princeton, 1996.
[Kristian and Sachs, 1966] J. Kristian and R. K. Sachs. Observations in cosmology. *Astrophysical Journal*, 143:379, 1966.
[Kuhn, 1977] T. Kuhn. Objectivity, value judgment, and theory choice. In *The Essential Tension*, pages 320–39. University of Chicago Pressm 1977.
[Lachièze *et al.*, 1995] M. M. R. Lachièze and J. P. Luminet. Cosmic topology. *Physics Reports*, 254:135, 1995.
[Lakatos, 1980] I. Lakatos. *The Methodology of Scientific Research Programmes*. Cambridge University Press, 1980.
[Landsman, 2006] N. P. Landsman. this vol, ch.5, 2006.
[Langlois and Vernizzi, 2005] D. Langlois and F. Vernizzi. Conserved non-linear quantities in cosmology. astro-ph/0509078, 2005.
[Lemaître, 1931] G. Lemaître. The beginning of the world from the point of view of quantum theory. *Nature*, 127:706, 1931.
[Leslie, 1989] J. Leslie. *Universes*. Routledge, 1989.
[Leslie, 1994] J. Leslie. Cosmology: a philosophical survey. *Philosophia*, 24:3–27, 1994.
[Leslie, 1998] J. Leslie, editor. *Modern Cosmology and Philosophy*. Prometheus Books, 1998.
[Lewis, 1986] D. K. Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford, 1986.
[Liddle, 2004] A. R. Liddle. How many cosmological parameters? astro-ph/0401198, 2004.
[Lidsey *et al.*, 1997] J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barriero, and M. Abney. Reconstructing the inflation potential—an overview. *Reviews of Modern Physics*, 69,:373–410, 1997.
[Linde, 1990] A. D. Linde. *Particle Physics and Inflationary Cosmology*. Harwood Academic, 1990.

[Lineweaver and Davis, 2005] C. H. Lineweaver and T. M. Davis. Misconceptions about the big bang. *Scientific American*, 2005.

[Luminet, 2005] J.-P. Luminet. A cosmic hall of mirrors. *Physics World*, 18:23–28, 2005.

[Luminet *et al.*, 2003] J.-P. Luminet, J. R. Weeks, A. Raizuelo, R. Lehoucq, and J. P. Uzan. Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background. *Nature*, 425:593–595, 2003.

[Madsen and Ellis, 1988] M. S. Madsen and G. F. R. Ellis. Evolution of Ω in inflationary universes. *Monthly Notice of the Royal Astronomical Society*, 234:67–77, 1988.

[Malament, 2006] D. Malament. This vol ch. 3, 2006.

[Matravers *et al.*, 1995] D. R. Matravers, G. F. R. Ellis, and W. R. Stoeger. Complementary approaches to cosmology: Relating theory and observations. *Quarterly Journal of the Royal Astronomical Society*, 36:29–45, 1995.

[McCrea, 1953] W. H. McCrea. Cosmology. *Reports on Progress in Physics*, 16:321–363, 1953.

[McCrea, 1960] W. H. McCrea. The interpretation of cosmology. *Nature*, 186:1035, 1960.

[McCrea, 1970] W. H. McCrea. A philosophy for big bang cosmology. *Nature*, 228:21, 1970.

[Meyer, 1994] D. M. Meyer A distant space thermometer. *Nature*, 371:13, 1994.

[Misner, 1968] C. W. Misner. The isotropy of the universe. *Astrophysical Journal*, 151:431–458, 1968.

[Misner, 1969] C. W. Misner. Mixmaster universe. *Physical Review Letters*, 22:1071–1074, 1969.

[Morowitz, 2002] L. Morowitz. *The emergence of everything: How the world came to be complex.* Oxford, 2002.

[Mouri and Taniguchi, 2005] H. Mouri and Y. Taniguchi. Hierarchical object formation in th epeculiar velocity field. astro-ph/0508069, 2005.

[Mulryne *et al.*, 2005] D. J. Mulryne, R. Tavakol, J. E. Lidsey, and G. F. R. Ellis. An emergent universe from a loop. *Physical Review D*, 71:123512, 2005.

[Munitz, 1962] M. K. Munitz. The logic of cosmology. *British Journal for the Philosophy of Science*, 13:34–50, 1962.

[Mustapha *et al.*, 1998] N. Mustapha, C. Hellaby, and G. F. R. Ellis. Large scale inhomogeneity vs source evolution: Can we distinguish them? *Monthly Notice of the Royal Astronomical Society*, 292:817–830, 1998.

[Nilsson *et al.*, 1991] J. S. Nilsson *et al.*, editors. *The Birth and Early Evolution of our Universe: Proceedings of Nobel Symposium 19.* World Scientific, 1991.

[North, 1965] J. D. North. *The Meaure of the Universe.* Oxford University Press, 1965.

[Padmanbhan, 1993] T. Padmanbhan. *Structure formation in the universe.* Cambridge University Press, 1993.

[Partridge, 1995] B. Partridge. *3K: The Cosmic Microwave Background Radiation.* Cambridge University Press, 1995.

[Peacock, 1999] J. A. Peacock. *Cosmological Physics.* Cambridge University Press, 1999.

[Peebles, 1971] P. J. E. Peebles. *Physical Cosmology.* Princeton University Press, Princeton, 1971.

[Peebles, 1993a] P. J. E. Peebles. *The Large-Scale Structure of the Universe.* Princeton University Press, 1993.

[Peebles, 1993b] P. J. E. Peebles. *Principles of Physical Cosmology.* Princeton University Press, 1993.

[Peebles *et al.*, 1991] P. J. E. Peebles, D. N. Schramm, E. L. Turner, and R. G. Kron. The case for the relativistic hot big bang cosmology. *Nature*, 352:769–776, 1991.

[Penrose, 1989a] R. Penrose. Difficulties with inflationary cosmology. In Fergus, E., editor, *14th Texas Symposium Relativistic Astrophysics.* New York Academy of Science, New York, 1989.

[Penrose, 1989b] R. Penrose. *The Emperor's New Mind.* Oxford University Press, 1989.

[Penrose, 2004] R. Penrose. *The Road to Reality.* Jonathan Cape, 2004.

[Perez *et al.*, 2005] A. Perez, H. Sahlmann, and D. Sudarsky. On the quantum origin of the seeds of cosmic structure. gr-qc/0508100, 2005.

[Perkins, 2005] D. Perkins. *Particle Astrophysics.* Oxford, 2005.

[Perlmutter *et al.*, 1998] S. Perlmutter *et al.* Discovery of a supernova explosion at half the age of the universe. *Nature*, 391:51, 1998.

[Pettini, 1999] M. Pettini. Element abundances at high redshifts. astro-ph/9902173, 1999.

[Rees, 1995] M. J. Rees. *Perspectives in Astrophysical Cosmology.* Cambridge University Press, 1995.

[Rees, 1999] M. J. Rees. *Just six numbers: The Deep Forces that shape the universe*. Weidenfeld and Nicholson, London, 1999.

[Rees, 2003] M. J. Rees. *Our Cosmic Habitat*. Princeton University Press, Princeton, 2003.

[Rindler, 1956] W. Rindler. Visual horizons in world models. *Monthly Notice of the Royal Astronomical Society*, 116:662, 1956.

[Rindler, 2001] W. Rindler. *Relativity: Special, General, and Cosmological*. Oxford University Press, 2001.

[Robertson, 1933] H. P. Robertson. Relativistic cosmology. *Reviews of Modern Physics*, 5:62–90, 1933.

[Robertson, 1935] H. P. Robertson. Kinematics and world structure. *Astrophysical Journal*, 82:284–301, 1935.

[Rothman and Ellis, 1993] T. Rothman and G. F. R. Ellis. Lost horizons cosmology. *American Journal of Physics*, 61:883–893, 1993.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity*. Cambridge University Press, 2004.

[Rovelli, 2006] C. Rovelli. This vol ch.12, 2006.

[Rugh and Zinkernagel, 2002] S. E. Rugh and H. Zinkernagel. The quantum vacuum and the cosmological constant problem. *Studies in History and Philosophy of Modern Physics*, 33:663–705, 2002.

[Saini *et al.*, 2000] T. D. Saini, S. Raychaudhury, V. Sahni, and A. A. Starobinsky. Reconstructing the cosmic equation of state from supernova distances. *Physical Review Letters*, 85:1162–1165, 2000.

[Sandage, 1961] A. Sandage. The ability of the 200-inch telescope to distinguish between selected world models. *Astrophysical Journal*, 133:355–392, 1961.

[Sandage *et al.*, 1993] A. Sandage, R. G. Kron, and M. S. Longair. *The Deep Universe: Saas-Fee Advanced Course 23*. Springer, Berlin, 1993.

[Schneider *et al.*, 1992] P. Schneider, J. Ehlers, and E. E. Falco. *Gravitational Lenses*. Springer, Berlin, 1992.

[Schramm and Turner, 1998] D. N. Schramm and M. S. Turner. Big-bang nucleosynthesis enters the precision era. *Reviews of Modern Physics*, 70:303–318, 1998.

[Schrödinger, 1956] E. Schrödinger. *Expanding Universes*. Cambridge University Press, Cambridge, 1956.

[Sciama, 1971] D. W. Sciama. Astrophysical cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology*, pages 183–236. Academic Press, New York, 1971.

[Scott, 2005] D. Scott. The standard cosmological model. astro-ph/0510731, 2005.

[Seife, 2000] C. Seife. *Zero: the Biography of a Dangerous Idea*. Penguin, 2000.

[Seife, 2003] C. Seife. *Alpha and Omega: The Search for the Begining and end of the Universe*. Bantam Books, London, 2003.

[Seife, 2004] C. Seife. Physics enters the twilight zone. *Science*, 305:464–466, 2004.

[Sievers *et al.*, 2005b] J. L. Sievers *et al.* Implications of the cosmic background imager polarization data, 2005. astro-ph/0509203.

[Sigurdson and Furlanetto, 2005] K. Sigurdson and S. R. Furlanetto. Measuring the primordial deuterium abundance during the cosmic dark ages. astro-ph/0505173.

[Silk, 1997] J. Silk. *A Short History of the Universe*. Scientific American Library, 1997.

[Silk, 2001] J. Silk. *The Big Bang*. Freeman, 3 edition, 2001.

[Silk, 2005] J. Silk. *On the shores of the unknown: A short history of the universe*. Cambridge University Press, 2005.

[Smith, 1982] R. W. Smith. *The Expanding Universe: Astronomy's Great Debate*. Cambridge University Press, 1982.

[Smolin, 1992] L. Smolin. Did the universe evolve? *Classical and Quantum Gravity*, 9:173–191, 1992.

[Spergel *et al.*, 2003] D. N. Spergel *et al.* First year Wilkinson microwave anisotropy probe (WMAP) observations: Determination of cosmological parameters. *Astrophysical Journal Supplement Series*, 148:175–194, 2003. astro-ph/0302209.

[Starkman and Schwarz, 2005] G. D. Starkman and D. J. Schwarz. Is the universe out of tune? *Scientific American*, pages 48–55, 2005.

[Steinhardt, 1995] P. J. Steinhardt. Cosmology confronts the cosmic microwave background. *International Journal of Modern Physics*, 10:1091–1124, 1995. astro-ph/0407329.

[Stoeger *et al.*, 1995] W. Stoeger, R. Maartens, and G. F. R. Ellis. Proving almost-homogeneity of the universe: an almost-Ehlers, Geren and Sachs theorem. *Astrophysical Journal*, 443:1–5, 1995. gr-qc/0508100.

[Stoeger *et al.*, 2004] W. R. Stoeger, G. F. R. Ellis, and U. Kirchner. Multiverses and cosmology: Philosophical issues. astro-ph/0407329, 2004.

[Susskind, 2003] L. Susskind. The anthropic landscape of string theory. hep-th/0302219, 2003.

[Susskind, 2005] L. Susskind. *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design*. Little Brown, New York, 2005.

[Susskind and Lindesay, 2004] L. Susskind and J. Lindesay. *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe*. World Scientific, 2004.

[Tegmark, 1998] M. Tegmark. Is 'the theory of everything' merely the ultimate ensemble theory? *Annals of Physics*, 270:1–51, 1998.

[Tegmark, 2002] M. Tegmark. Measuring spacetime: from big bang to blackholes. *Science*, 296:1427–1433, 2002. astro-ph/0207188.

[Tegmark, 2003] M. Tegmark. Parallel universes. In J. D. Barrow, P. C. W. D. and Harper, C., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity*, pages 459–491. Cambridge University Press, 2003. astro=ph/0302131.

[Tegmark *et al.*, 2004] M. Tegmark (the SDSS collaboration). Cosmological parameters from SDSS and WMAP. *Physical Review D*, 69:103501, 2004.

[Tegmark *et al.*, 2003] M. Tegmark, A. Vilenkin and L. Pogosan. Anthropic predictions for neuturinos masses. astro-ph/0304536, 2003.

[Tipler *et al.*, 1980] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis. Singularities and horizons: A review article. In Held, A., editor, *General Relativityand Gravitation*. Plenum Press, 1980.

[Tolman, 1934] R. C. Tolman. *Relativity, Thermodynamics, and Cosmology*. Clarendon Press, Oxford, 1934. Dover Publications, 1987.

[Tryon, 1973] E. P. Tryon. Is the universe a quantum fluctuation? *Nature*, 246:396, 1973.

[Uggla *et al.*, 2003] C. Uggla, H. van Elst, J. Wainwright, and G. F. R. Ellis. The past attractor in inhomogeneous cosmology. *Phys Rev.*, D68:103502 (1-22), 2003.

[Uffink, 2006] J. Uffink. This vol Ch 9, 2006.

[Vilenkin, 1982] A. Vilenkin. Creation of universes from nothing. *Physics Letters B*, 117:25, 1982.

[Wainright *et al.*, 1998] J. Wainright, A. A. Coley, G. F. R. Ellis, and M. Hancock. On the isotropy of the universe: Do Bianchi VIIh universes isotropize? *Classical and Quantum Gravity*, 15:331, 1998.

[Wainwright and Ellis, 1996] J. Wainwright, and G. F. R. Ellis. *The dynamical systems approach to cosmology*. Cambridge University Press, 1996.

[Wald, 2005] R. M. Wald. The arrow of time and the initial conditions for the universe. gr-qc/0507094, 2005.

[Wald, 1983] R. M. Wald. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. *Physical Review*, 28:2118, 1983.

[Walker, 1936] A. G. Walker. On Milne's theory of world structure. *Proceedings of the London Mathematical Society*, 42:90, 1936.

[Walker, 1944] A. G. Walker. Completely symetric spaces. *Journal of the London Mathematical Society*, 19:219–226, 1944.

[Weeks *et al.*, 2003] J. Weeks, J.-P. Luminet, A. Riazuelo, and R. Lehoucq. Well-proportioned universes suppress CMB quadrupole. astro-ph/0312312, 2003.

[Weinberg, 1989] S. W. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23, 1989.

[Weinberg, 1972] S. W. Weinberg. *Gravitation and Cosmology*. Wiley, New York, 1972.

[Weinberg, 2000a] S. W. Weinberg. The cosmological constant problem. astro-ph/0005265, 2000.

[Weinberg, 2000b] S. W. Weinberg. A priori distribution of the cosmological constant. *Physical Review D*, 61:103505. astro-ph/0002387, 2000.

[Wheeler, 1968] J. A. Wheeler. *Einstein's Vision*. Springer, Berlin, 1968.

[White *et al.*, 1993] M. White, L. M. Krauss, and J. Silk. Cosmic variance in CMB anisotropies: From 1° to COBE. *Astrophysical Journal*, 418:535–543, 1993.

[Will, 1979] C. M. Will. The confrontation between gravitational theory and experiment. In Hawking, S. W. and Israel, W., editors, *General Relativity: an Einstein Centenary Survey*, pages 24–89. Cambridge University Press, 1979.

[Wilson, 1989] C. Wilson. *Leibniz's Metaphysics: A Historical and Comparative Study*. Princeton University Press, 1989.

[Wright, 2006] E. L. Wright. A century of cosmology. Astro-ph/0603750.

[Yoshimura, 1988] M. Yoshimura. The universe as a laboratory for high energy physics. In Fang, L.-Z. and Zee, A., editors, *Cosmology and Particle Physics*, pages 293–340. Gordon and Breach, 1988.

[Zeh, 1992] H. D. Zeh. *The Physical Basis of the direction of Time*. Springer Verlag, Berlin, 1992.

[Zinkernagel, 2002] H. Zinkernagel. Cosmology, particles, and the unity of science. *Studies in History and Philosophy of Modern Physics*, 33:493–516, 2002.

Issues in the Philosophy of Cosmology

SUMMARY TABLE OF ISSUES AND THESES

Issue A: The uniqueness of the universe

Thesis A1: The universe itself cannot be subjected to physical experimentation

Thesis A2: The universe cannot be observationally compared with other universes

Thesis A3: The concept of 'Laws of Physics' that apply to only one object is questionable

Thesis A4: The concept of probability is problematic in the context of existence of only one object

Issue B: The large scale of the Universe in space and time

Thesis B1: Astronomical observations are confined to the past null cone, and fade with distance

Thesis B2: 'Geological' type observations can probe the distant past of our past world line

Thesis B3: Establishing a Robertson-Walker geometry relies on plausible philosophical assumptions

Thesis B4: Interpreting cosmological observations depends on astrophysical understanding

Thesis B5: A key test for cosmology is that the age of the universe must be greater than the ages of stars

Thesis B6: Horizons limit our ability to observationally determine the very large scale geometry of the universe

Thesis B7: We have made great progress towards observational completeness

Issue C: The unbound energies in the early universe

Thesis C1: The Physics Horizon limits our knowledge of physics relevant to the very early universe

Thesis C2: The unknown nature of the inflaton means inflationary universe proposals are incomplete

Issue D: Explaining the universe — the question of origins

Thesis D1: An initial singularity may or may not have occurred

Thesis D2: Testable physics cannot explain the initial state and hence specific nature of the universe

Thesis D3: The initial state of the universe may have been special or general

Issue E: The Universe as the background for existence

Thesis E1: Physical laws may depend on the nature of the universe

Thesis E2: We cannot take the nature of the laws of physics for granted

Thesis E3: Physical novelty emerges in the expanding universe

Issue F: The explicit philosophical basis

Thesis F1: Philosophical choices necessarily underly cosmological theory

Thesis F2: Criteria for choice between theories cannot be scientifically chosen or validated

Thesis F3: Conflicts will inevitably arise in applying criteria for satisfactory theories

Thesis F4: The physical reason for believing in inflation is its explanatory power re structure growth.

Thesis F5: Cosmological theory can have a wide or narrow scope of enquiry

Thesis F6: Reality is not fully reflected in either observations or theoretical models

Issue G: The Anthropic question: fine tuning for life

Thesis G1: Life is possible because both the laws of physics and initial conditions have a very special nature

Thesis G2: Metaphysical uncertainty remains about ultimate causation in cosmology

Issue H: The possible existence of multiverses

Thesis H1: The Multiverse proposal is unprovable by observation or experiment

Thesis H2: Probability-based arguments cannot demonstrate the existence of multiverses

Thesis H3: Multiverses are a philosophical rather than scientific proposal

Thesis H4: The underlying physics paradigm of cosmology could be extended to include biological insights

Issue I: The natures of existence

Thesis I1: We do not understand the dominant dynamical matter components of the universe at early or late times

Thesis I2: The often claimed physical existence of infinities is questionable

Thesis I3: A deep issue underlying the nature of cosmology is the nature of the laws of physics.

Thesis of Uncertainty: Ultimate uncertainty is one of the key aspects of cosmology