10.1098/rsta.2003.1288

THE ROYAL SOCIETY

A quintessential introduction to dark energy

By Paul J. Steinhardt

Department of Physics, Princeton University, Princeton, NJ 08540, USA

Published online 17 September 2003

Most of the energy in the Universe consists of some form of dark energy that is gravitationally self-repulsive and that is causing the expansion of the Universe to accelerate. The possible candidates are a vacuum energy density (or, equivalently, a cosmological constant) and quintessence, a time-evolving, spatially inhomogeneous component with negative pressure. In this review, we focus on quintessence and ideas on how it might solve the cosmic coincidence problem, how it might be distinguished observationally from a cosmological constant, and how it may affect the overall cosmic history of the Universe.

Keywords: quintessence; *k*-essence; dark energy; cyclic models; microwave background; supernovae

1. Introduction

The discovery of dark energy is one of the most surprising and profound discoveries in the history of science. Some of its implications are the following.

Most of the energy in the Universe is *not* 'matter'. In its first 300 years, physics has focused on the properties of matter and radiation, including dark matter. Now we know that they represent less than 30% of the composition of the Universe. The rest consists of something we know virtually nothing about.

Most of the energy in the Universe is *not* gravitationally attractive. We are probably the last generation to have been taught that 'gravity always attracts', a notion which has been presented as a basic fact of nature for hundreds of years. We are now aware that gravity can repel, as well. Of course, the possibility of self-repulsive forms of energy was there in Einstein's general theory of relativity since its inception, but this point has not generally been appreciated until now. We must rewrite the textbooks to explain that the gravitationally self-attracting matter with which we are familiar is the *minority* in the Universe today (and for the indefinite future).

We live at a special time in the history of the Universe. The Copernican revolution taught us that there is nothing special about our location in the Universe. If space is uniform, then should not the same be true for time? Hubble's discovery that the Universe is expanding taught us that the Universe is evolving,

One contribution of 13 to a Discussion Meeting 'The search for dark matter and dark energy in the Universe'.

but the notion was that the evolution has been steady over the last 15 billion years with no remarkable changes. We now know that time is anti-Copernican. We live at a special moment in cosmic history: the transition between a decelerating, matter-dominated Universe and an accelerating, dark-energy-dominated Universe. The progressive formation of ever-larger scale structure and increasing complexity that characterized the matter-dominated Universe has reached an end, and now the Universe is headed towards a period that is ever-emptier and structureless.

The future (and perhaps the past) is determined by dark energy. The immediate future of the Universe will be governed by dark energy, which will determine the rate of dilution and cooling of the matter and energy. But, perhaps dark energy plays a more profound role in the history of the Universe, determining our distant past as well as our long-term future. We will discuss the recent proposal of a 'cyclic Universe' (Steinhardt & Turok 2002*a*), in which dark energy is a key part of the engine driving the periodic evolution of the Universe.

Given the profound implications above, finding the identity of dark matter has emerged as one of the most important scientific challenges of the twenty-first century. The first evidence for dark energy emerged in the mid 1990s (Ostriker & Steinhardt 1995; Krauss & Turner 1995). First, improved observations confirmed that the total mass density is probably less than half of the critical density (Bahcall *et al*. 1995; Bahcall & Fan 1998; Carlberg *et al*. 1996). At the same time, combined measurements of the cosmic-microwave-background (CMB) temperature fluctuations and the distribution of galaxies on large scales began to suggest that the Universe is flat, consistent with the standard inflationary prediction. The only way to have low mass density and a flat Universe, as expected from the inflationary theory, is if an additional, non-luminous, 'dark' energy component dominates the Universe today. The dark energy would have to resist gravitational collapse, or else it would already have been detected as part of the clustered energy in the haloes of galaxies. But, as long as most of the energy of the Universe resists gravitational collapse, it is impossible for structure to form in the Universe. The dilemma can only be resolved if the hypothetical dark energy was negligible in the past and, then, only after galaxies and larger scale structure formed, it became the dominant energy in the Universe. According to general relativity, the only type of energy with this property has *negative* pressure. This simple argument (Ostriker & Steinhardt 1995) rules out almost all of the usual suspects, such as cold dark matter, neutrinos, radiation and kinetic energy, because they have zero or positive pressure. Furthermore, according to Einstein's equations, negative pressure implies cosmic acceleration. So, this analysis anticipated the supernovae results (Perlmutter *et al*. 1998; Riess *et al*. 1998) which have provided direct evidence for acceleration.

Hence, there are numerous lines of evidence establishing that dark energy exists and that it comprises nearly 70% of the energy density of the Universe today. But, all of these observations do little to inform us about what the dark energy is.

2. What is the dark energy?

The two logical possibilities for dark energy are the cosmological constant and quintessence. The cosmological constant was first introduced by Einstein for the purpose of constructing a static model of the Universe. The repulsive cosmological constant was delicately fine-tuned to balance the gravitational attraction of matter (Einstein 1917). Today, the cosmological constant is recognized as vacuum energy, an energy assigned to empty space itself that has negative pressure and induces cosmic acceleration. It has the same value everywhere in space for all time, and it is chemically inert. And, unlike Einstein's original concept, the cosmological constant, if it comprises the dark energy, has not been fine-tuned to balance the matter. Instead, the vacuum energy is overabundant, causing the expansion of the Universe to accelerate. The cosmological constant is completely defined by one number, its magnitude.

Quintessence is a dynamical, evolving, spatially inhomogeneous component with negative pressure (Caldwell *et al.* 1998). The term derives from the ancient word for 'fifth element'; according to some mediaeval metaphysicians, the Universe consists of earth, air, fire and water, plus an additional all-pervasive, component that accounted for the motion of the Moon and planets. By analogy, in the current context, quintessence would be the fifth dynamical component that has influenced the evolution of the Universe, in addition to the previously known baryons, leptons, photons and dark matter.

Quintessence is characterized by its equation of state $w \equiv p/\rho$, where p is the pressure and ρ is the energy density. Most models have $0 \geqslant w > -1$, whereas a cosmological constant has w precisely equal to -1. The smaller is the value of w, the greater its accelerating effect. Unlike a cosmological constant, the quintessential pressure and energy density evolve in time, and w may also do so. Furthermore, because the quintessence component evolves in time, it is, by general covariance, necessarily spatially inhomogeneous. In some models, quintessence also has a time-varying speed of sound that can enhance the effect of fluctuations on the CMB and large-scale structure.

It should be emphasized that the quintessence explanation for the dark energy does not explain the longstanding problem of why the cosmological constant is so small compared with the Planck scale. Prior to the discovery of dark energy, it had been presumed that some symmetry or cancellation mechanism causes the vacuum energy to vanish altogether or to reach a level where it is negligibly small. If the dark energy proves to be quintessence, we would need to invoke the same cancellation mechanisms.

A common model of quintessence is the energy density associated with a scalar field Q slowly rolling down a potential $V(Q)$. The pressure of the scalar field,

$$p = \tfrac{1}{2}\dot{Q}^2 - V(Q),$$

is negative if the field rolls slowly enough that the kinetic energy density is less than the potential energy density. The ratio of kinetic energy to potential energy is determined by the equation of motion for the scalar field:

$$\ddot{Q} + 3H\dot{Q} + V'(Q) = 0. \tag{2.1}$$

This determines the equation of state

$$w_Q \equiv \frac{p}{\rho} = \frac{\tfrac{1}{2}\dot{Q}^2 - V(Q)}{\tfrac{1}{2}\dot{Q}^2 + V(Q)}. \tag{2.2}$$

For most potentials, w evolves slowly with time. The field is assumed to couple only gravitationally to matter. The Q-energy density decreases with time as $1/a^{3(1+w_Q)}$, so negative pressure corresponds to a density which decreases more slowly than $1/a^3$.

The spatial inhomogeneities in Q evolve over time due to the gravitational interaction between Q and clustering matter (Caldwell *et al.* 1998). The perturbations are important because they can leave a distinguishable imprint on the CMB and large-scale structure. To determine how the perturbations evolve, specifying w is insufficient. One must know the response of the quintessence component to perturbations. This can be defined by specifying the sound speed c_{s} as a function of wavenumber k or, alternatively, by specifying the microphysics and equations of motion (from which the perturbative equations can be derived). Note that it is possible, in principle, to have two fluids with the same w but different c_{s}, which would lead to distinct observational predictions, as discussed later in this paper.

For a scalar field, the equation of motion for the perturbations δQ in synchronous gauge is

$$\delta\ddot{Q} + 3H\delta\dot{Q} + (c_{\mathrm{s}}^2 k^2 + a^2 V''(Q))\delta_Q = -\tfrac{1}{2}\dot{h}_k\dot{Q}, \tag{2.3}$$

where the dot represents the derivative with respect to conformal time, the prime represents the derivative with respect to Q, and h_k is the kth fourier mode of the perturbed metric. Consider the source term in equation (2.3). First, any realistic cosmological model includes clustering matter components (baryons and dark matter), so $\dot{h}_k$ must be non-zero. Also, $\dot{Q}$ is non-zero. Hence, the source term on the right-hand side must be non-zero overall. This is significant, because it ensures that Q cannot be smoothly spread. Even if δQ is zero initially, the source term ensures that perturbations to grow.

A further consequence of the source term is that the perturbations in Q observed today are extremely insensitive to the initial conditions for δQ (Caldwell *et al.* 1998). Assuming that $\delta\rho_Q/\rho_Q$ is comparable with the perturbations in other energy components, the transient solution to equation (2.3) is negligible today compared with the particular solution set by the source term.

Why consider quintessence if its effect on the expansion of the Universe is similar to the cosmological constant? The principle reasons are

(i) quintessence has different implications for fundamental physics;

(ii) quintessence may explain the 'cosmic coincidence' problem;

(iii) quintessence may fit the observational data better than the cosmological constant; and

(iv) quintessence may suggest a radically new picture of the overall history of the Universe.

The first point is clear: whatever its identity, dark energy must be now incorporated in any future attempt at a unified theory of fundamental interactions. A vacuum density or cosmological constant (Λ) is static and spatially uniform. Its value is set once and for all in the very early Universe and is tied directly to quantum gravity physics near the Planck scale. Quintessence is new dynamics at ultra-low energies (energy scale *ca.* 1 meV today), perhaps a harbinger of a whole spectrum of new low-energy phenomena. In addition, for quintessence there is the added observational constraint that its coupling to ordinary matter be sufficiently suppressed to evade fifth force and other constraints on light fields (Carroll 1998).

The last three points are addressed in each of the remaining sections.

3. Fine-tuning, cosmic coincidence, and the quintessential solution

Whatever form the dark energy takes, two new cosmological problems arise. First, the component must have a tiny energy density (*ca.* $10^{-47}\,\mathrm{GeV}^4$) today. How does this small value arise from a microphysical theory? We will refer to this puzzle as the 'fine-tuning problem'.

A second problem arises when the cosmological model is extrapolated back in time to the very early Universe, at the end of inflation, say. The quintessence energy density decreases at a different rate from the matter density, and their ratio shrinks by many orders of magnitude as we extrapolate back in time. The observations tell us that, somehow, the ratio was set initially just right so that now, fifteen billion years later, the ratio is of order unity. Accounting for the special ratio in the early Universe will be referred to as the 'coincidence problem' (Steinhardt 1997). The coincidence problem is a generalization of the flatness problem pointed out by Dicke & Peebles (1979).

The fine-tuning and cosmic coincidence problems are vexing. They are often posed as a paradox: why should the acceleration begin just as humans evolve? In desperation, some cosmologists and physicists have given renewed attention to anthropic models (Weinberg 2000). But many continue to seek a dynamical explanation which does not require the fine-tuning of initial conditions or mass parameters and which is decidedly non-anthropic. A dynamical approach would seem to demand some sort of quintessence solution, since it would have to entail some interaction between the dark energy and the matter–radiation background.

It might appear that replacing a cosmological constant with a scalar field and potential energy is a step backwards. First, a general potential will not do. There must be a value of Q such that $V(Q)$ equals today's dark energy density ($10^{-47}\,\mathrm{GeV}^4$). Second, we must explain why the field has that particular value today. In general, this is not simply a matter of choosing the potential, but also a matter of carefully choosing the initial value of the field and its time derivatives. So, instead of tuning one parameter, the cosmological constant, we must tune the parameters of the potential *and* the initial conditions in the field.

However, some creative solutions have been introduced to address the problem. We will focus here on a single example which has combines several of these creative concepts. The example, known as k-essence (Armendariz-Picon *et al*. 2000, 2001), is a form of quintessence model in which the action for the scalar field has purely kinetic terms and no potential terms. In addition to the canonical kinetic energy density term, $X \equiv \frac{1}{2}(\partial Q)^2$, the k-essence action has higher-order nonlinear kinetic couplings. The Lagrangian density can be written as

$$\mathcal{L} = -\tfrac{1}{6}R + \frac{1}{Q^2}\tilde{p}_k(X) + \mathcal{L}_m, \tag{3.1}$$

where R is the Ricci scalar, and $\mathcal{L}_m$ is the Lagrangian density for dust and radiation and we use units where $\frac{8}{3}\pi G = 1$. The energy density is $\rho_k = (2X\tilde{p}_{k,X} - \tilde{p}_k)/Q^2$; the pressure is $p_k = \tilde{p}_k/Q^2$; and the speed of sound of k-essence is $c_{\mathrm{s}}^2 = p_{k,X}/\rho_{k,X}$. In

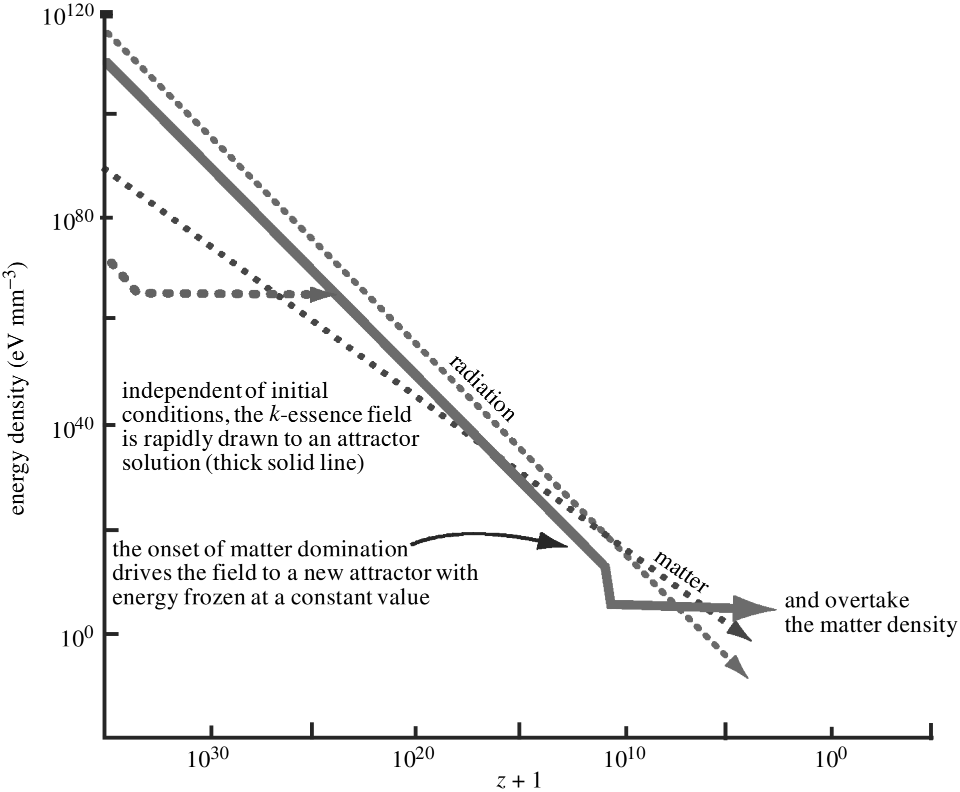

Figure 1. A plot showing the matter, radiation and quintessence energy density as a function of redshift for k-essence models. The k-essence models have dynamical attractor solutions that funnel a wide range of initial conditions (dashed) into a common evolutionary track (thick solid, upper left). At the onset of matter domination, the k-essence field switches to a new attractor solution that acts like a cosmological constant (bottom).

string and supergravity theories, non-standard kinetic terms appear generically in the effective action describing the massless scalar degrees of freedom. Normally, the nonlinear terms are ignored because they are presumed to be small and irrelevant. This is a reasonable expectation since the Hubble expansion damps the kinetic energy density over time. However, one case in which the nonlinear terms cannot be ignored is if there is an attractor solution which forces the nonlinear terms to remain non-negligible. This is precisely what occurs here. Hence, we wish to emphasize that k-essence models are constructed from building blocks that are common to most quantum-field theories. It is the dynamical attractor behaviour (that often arises in models with nonlinear kinetic energy) which is responsible for the highly novel features. The story is summarized in figure 1.

First, the nature of an attractor equation is that the evolution of the scalar field is completely insensitive to the initial value of the field and its time derivatives. As indicated in figure 1, the evolution of the dark energy component rapidly approaches an attractor solution which depends only on the action itself. What is more remarkable is that the attractor solution depends on what is happening in the rest of the Universe. If the Universe is radiation dominated, the k-essence behaves as if it were another radiation component with $w = 1/3$, and its energy density decreases in parallel with the dominant radiation component. Quintessence models with this property are called 'trackers', examples of which include models with non-zero potentials (Zlatev *et al.* 1998; Steinhardt *et al.* 1999). So, not only is the evolution independent of the initial conditions, but the tracking behaviour ensures that the energy density of the k-essence field remains negligible compared with the radiation density throughout the radiation-dominated epoch. We have a dynamical explanation for why the dark energy did not overtake the Universe for the first 10 000 years. But, then, something truly remarkable happens to k-essence models when the Universe becomes matter dominated. The radiation-like attractor solution becomes unstable, and the energy density in the k-essence field begins to drop several orders of magnitude until a new matter-dominated attractor solution is found. This attractor solution keeps the k-essence density constant (see figure 1), as if $w = -1$. The drop in energy density means that the k-essence cannot dominate immediately. But, once it hits the $w = -1$ attractor, the Universe can only expand for a short term before k-essence overtakes the Universe and throws it into a phase of cosmic acceleration.

In this scenario, the coincidence problem is beautifully addressed. Why did the Universe begin to accelerate just as humans started to evolve? Cosmic acceleration and human evolution are both linked to the onset of matter domination. The k-essence component has the property that it only behaves as a negative pressure component after matter–radiation equality, so that it can only overtake the matter density and induce cosmic acceleration after the matter has dominated the Universe for some period, at about the present epoch. And, of course, human evolution is linked to matter domination because the formation of planets, stars, galaxies and large-scale structure only occurs after the beginning of the matter-dominated epoch.

At this point, the understanding of k-essence models is rather primitive, and the worked examples are not very appealing in detail (too many terms and parameters). A better understanding of nonlinear attractor behaviour is needed to see if simple, plausible examples can be found. However, conceptually, k-essence is an important example of a dynamical, non-anthropic explanation of the fine-tuning and cosmic coincidence problems that might arise from a fundamental theory.

4. Distinguishing quintessence from the cosmological constant

Distinguishing quintessence from the cosmological constant is a difficult challenge. We must take advantage of their subtle differences.

First, quintessence predicts a different value of w and, hence, a different acceleration rate from vacuum energy ($w = -1$). The effect is to change slightly the relation between angular or luminosity distance and redshift. In figure 2, we compare two models with identical cosmic parameters except that the value of w differs. Note that the position of the first acoustic peak changes systematically as w changes. Of course, even more apparent are the changes in the heights of the peaks. Unfortunately, neither effect can be used as a clear diagnostic for distinguishing quintessence from a cosmological constant or determining w. As pointed out by Huey *et al.* (1999), there is a degeneracy problem: a combination of variations in Ω_{m} (the ratio of the matter density to the critical density), the Hubble parameter, the curvature and w keep the

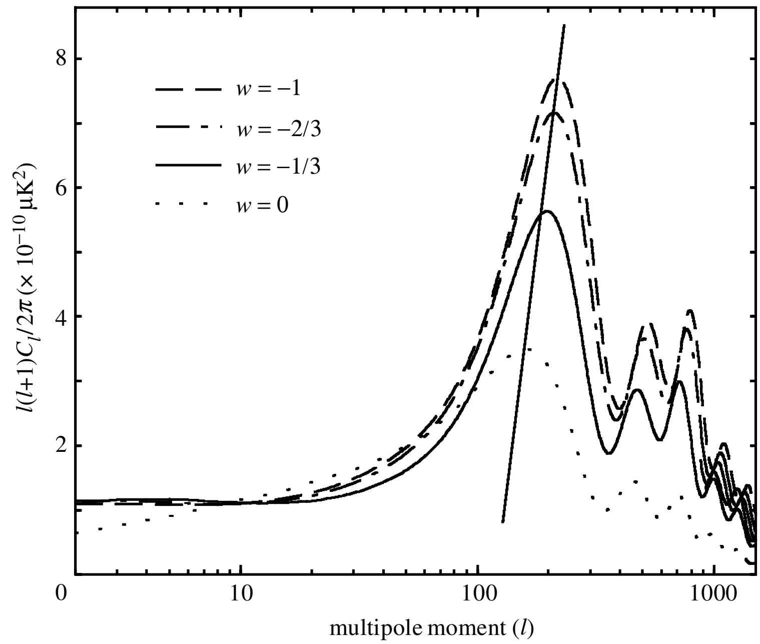

Figure 2. The CMB power spectrum (multipole moments C_ℓ versus multipole number ℓ) for a sequence of models with identical parameters except for varying w. Note the small shift in the position of the first acoustic peak as w varies, as indicated by the tilted line.

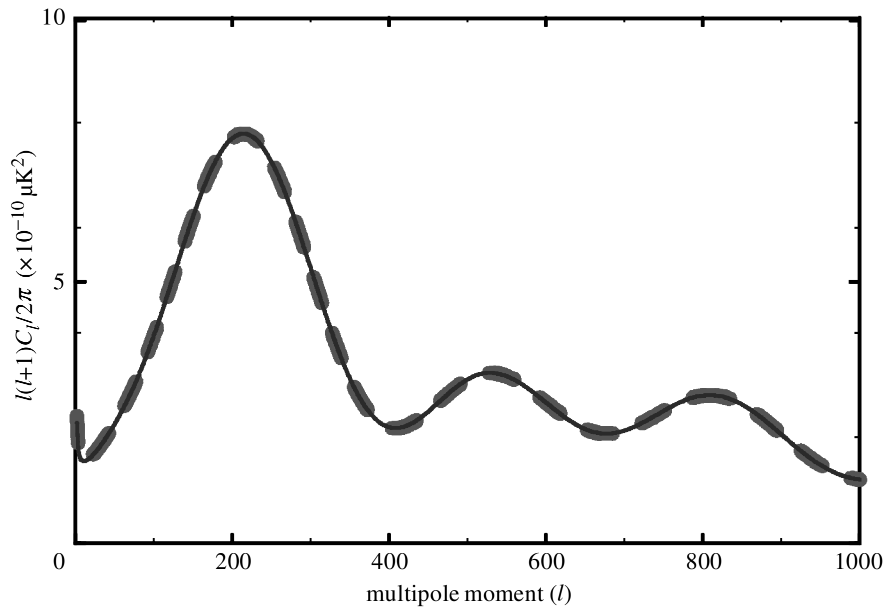

Figure 3. An illustration of the CMB degeneracy problem: the CMB power spectra completely overlap for a flat quintessence model with $w = -0.56$, $\Omega_{\rm m} = 0.30$, $\Omega_Q = 0.7$, and $h = 0.56$ (thick dashed line) and a closed model with $w = -1$, $\Omega_{\rm m} = 0.35$, $\Omega_\Lambda = 0.7$ and $h = 0.6$ (thin solid line).

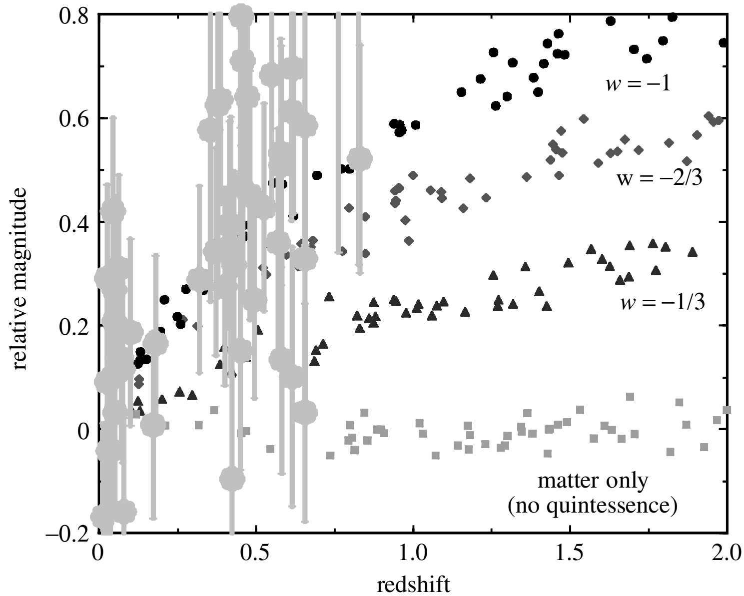

Figure 4. Magnitude versus redshift relative to a model with no quintessence ($\Omega_m = 1$) for $\Omega_m = 0.3$ and $w = -1$, $-2/3$, $-1/3$ and 0. Small symbols represent a simulation assuming a measurement of thousands of supernovae with redshift $z < 2$. Data points with error bars (on the left) represent current data from individual supernovae.

CMB power spectrum nearly unchanged for constant $w < -1/2$. Figure 3 shows a dramatic example. As a result, if w is greater than $-1/2$ or rapidly time-varying, then the microwave background only constrains a combination of parameters and one must use other tests to resolve w independently.

A way of constraining w and the acceleration directly is by measurements of type-IA supernovae at deep redshift. Figure 4 illustrates how well supernova measurements over a range of redshift $0 < z < 2$ can do in discriminating models with different constant w. The small symbols represent what can be obtained by measuring thousands of supernovae with optimal accuracy and using the systematic errors projected by the Supernova Acceleration Probe (SNAP) team. One obtains an optimistic impression of how well w can be resolved. However, caution is due. The points with the large error bars (towards the left) represent the first 40 supernovae that have been measured and their error bars. These are less impressive, offering virtually no discrimination in the most likely range, $-1 \leqslant w \leqslant -2/3$. The most optimistic projections rely on the assumption that the systematic errors are very small and that, by measuring thousands of supernovae, the statistical errors can be reduced to the size of the small symbols.

Even with this assumption, there remains an additional degeneracy problem that cannot be resolved. Namely, if we do not assume $w(z)$ is constant, but, instead, consider the possibility that w varies with redshift (as in k-essence models, for example), then the ability of supernovae surveys to resolve w today or its time-variation is enormously reduced. Figure 5 shows a group of models with widely varying w and $\mathrm{d}w/\mathrm{d}z$ today, along with the corresponding predictions for luminosity distance $d_L(z)$. The figure illustrates a fundamental degeneracy that makes it difficult to resolve w to

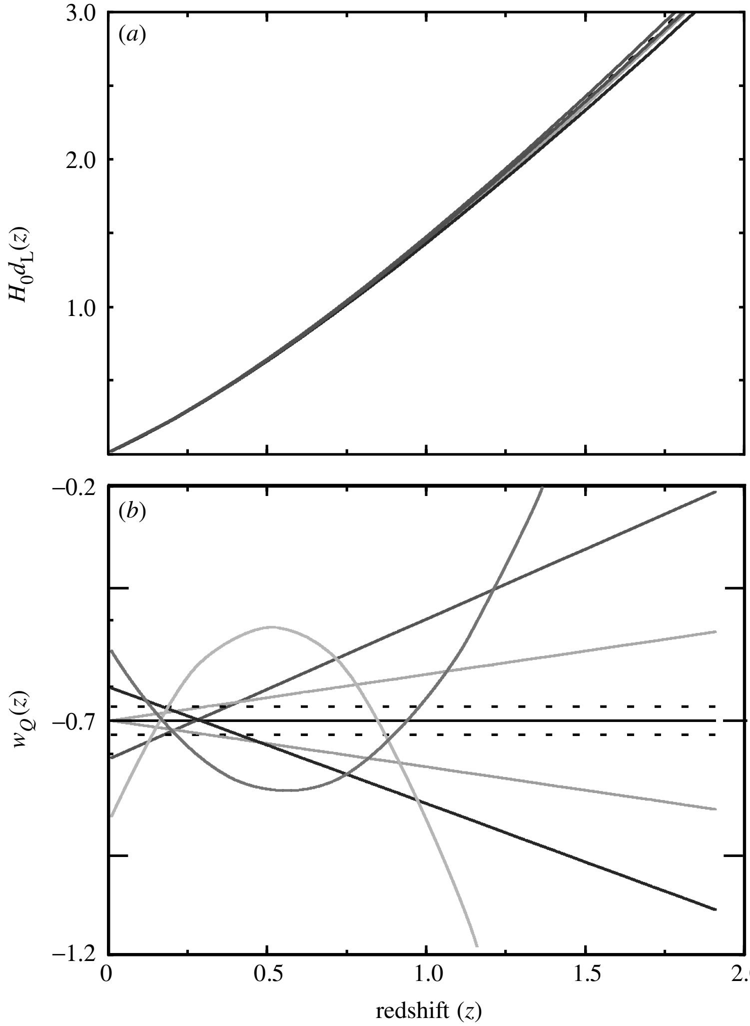

Figure 5. An illustration of the supernova degeneracy problem: the luminosity distance $d_L(z)$ versus z curves nearly overlap in (*a*) for the nine cases of $w(z)$ shown in (*b*). All models have $\Omega_m = 0.3$, and H_0 is the current value of the Hubble parameter.

much better than 40% or to obtain any useful information about dw/dz (Maor *et al.* 2001). The interested reader should consult Maor *et al.* (2002) to see more illustrations and details of this degeneracy problem.

At present, I am unaware of any probe or combination of probes that can precisely determine w and its time-variation (Maor *et al.* 2002). Many microphysical models predict that w is nearly constant, and currently planned tests can be useful for selecting out some of these possibilities. But a key challenge in the field is to find a better, more general test.

Another distinctive property of quintessence is that it is spatially inhomogeneous. Equation (2.3) can be used to predict the fluctuations in the quintessence energy density. The biggest effect is on the large-angular-scale microwave background

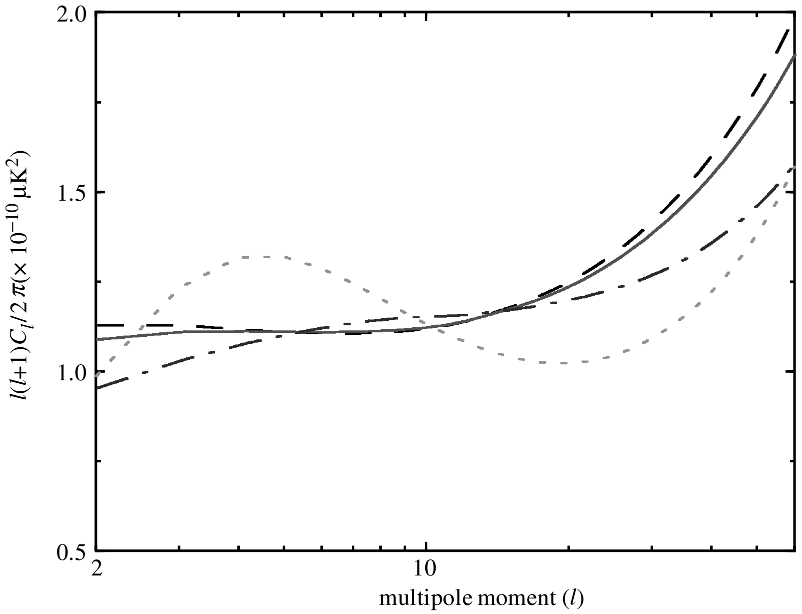

Figure 6. A comparison of the large-angular-scale microwave background anisotropy (lowest multipole moments) for a model with cosmological constant (dashed line), and quintessence models with $w = -2/3$ (solid line), $w = -1/3$ (dot-dashed line) and rapidly varying $w(z)$ with $w = 0$ at present (dotted line).

anisotropy, because the quintessence fluctuations are weak compared with the matter fluctuations at smaller scales and the quintessence energy density is negligible when those length-scale enter the horizon.

On large angular scales, quintessence fluctuations can alter the low-multipole moments of the CMB power spectrum. This modification is in addition to the usual (late) integrated Sachs–Wolfe (ISW) effect. The ISW occurs in any model with $\Omega_{\mathrm{m}} < 1$, whether an open model, a model with quintessence or a model with a cosmological constant. It comes about because the gravitational potential well due to a mass fluctuation changes as a CMB photon traverses the fluctuation passing from the last scattering surface to the present. The net ISW effect is to increase the multipole moments on angular scales which enter the horizon when $\Omega_{\mathrm{m}} < 1$, that is to say, the low-ℓ multipole moments. Fluctuations in the quintessence component cancel this effect because they add to the gravitational potential (Dave *et al.* 2002).

Even at the largest angular scales, the fluctuation effect is weak, becoming completely negligible as w approaches -1. Figure 6 shows the low-multipole moments (large-angular-scale anisotropy) of the microwave background temperature–power spectrum for models with constant and time-varying w. Unfortunately, the fluctuation effect is very small unless $w \geqslant -1/3$ or very rapidly time-varying, which is inconsistent with other cosmological constraints.

Another way to distinguish the nature of dark energy is to measure its sound speed to determine whether it is different from unity (the speed of light). The sound speed can be detected because it also affects the perturbations in the quintessence energy distribution. This approach is less generic because the sound speed in many

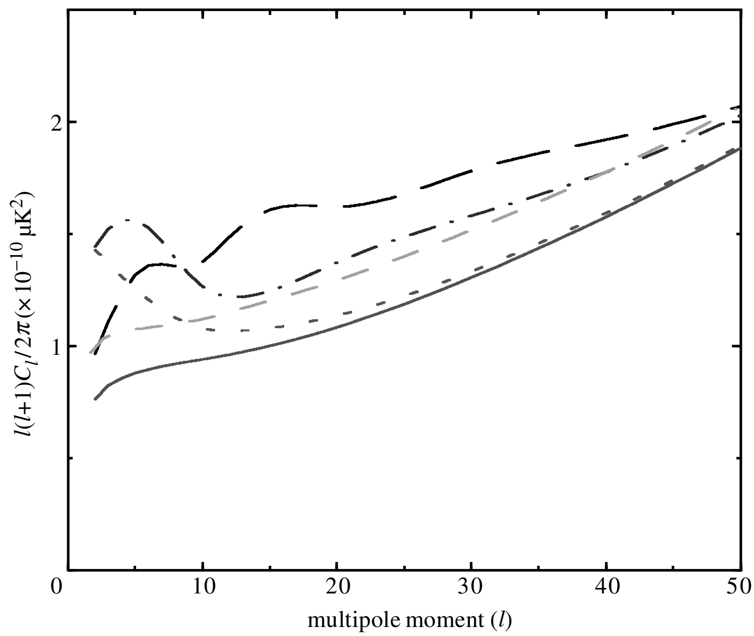

Figure 7. Comparison of the lowest multipole moments of the microwave background temperature power spectrum for a series of models with the same $w(z)$, but different sound speed $c_s(z)$: (*a*) $c_s = 1$ (dotted line); (*b*) $c_s = 1$ for $z > 10$ and $c_s = 0$ for $z < 10$ (solid line); and a sequence of k-essence models (dot-dashed, short- and long-dashed lines).

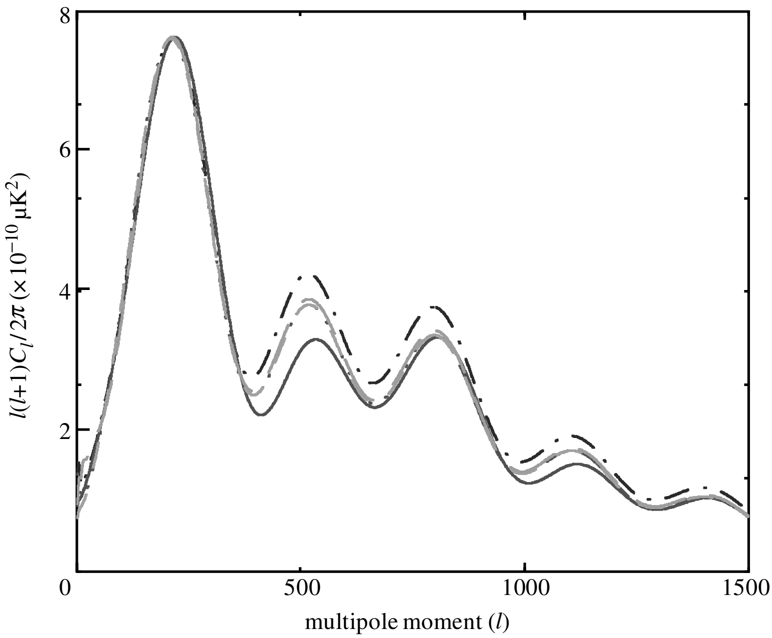

Figure 8. Comparison of higher multipole moments of the microwave background temperature power spectrum for the models in figure 7. The spectra have been normalized so that the amplitudes match at the top of the first acoustic peak.

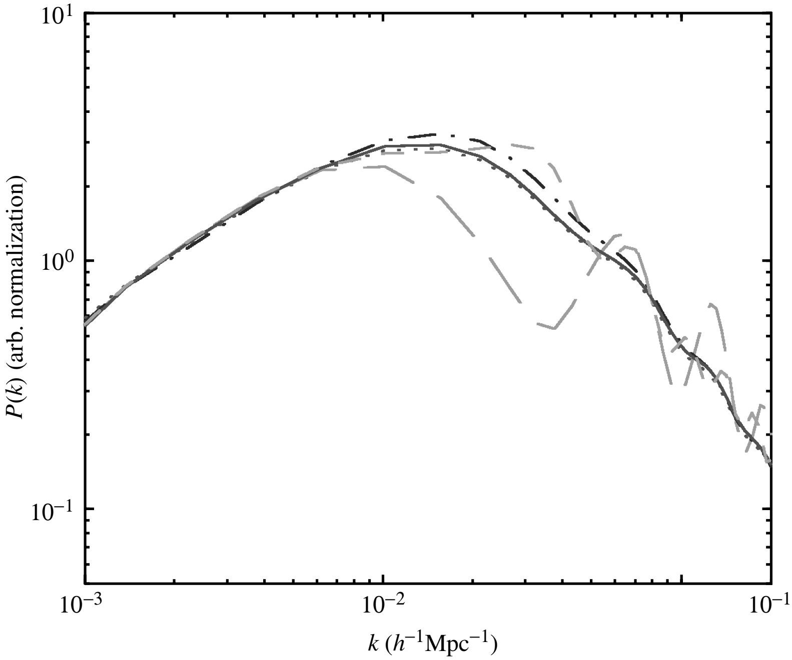

Figure 9. Comparison of the mass power spectra, $P(k)$ as a function of wavenumber k for the sequence of models in figure 7. The normalization of the curves is arbitrary.

models of quintessence is equal to unity, e.g. models in which quintessence consists of a scalar field (Q) with canonical kinetic energy density ($X \equiv \frac{1}{2}(\partial_\mu Q)^2$) and a positive potential energy density ($V(Q)$). However, in general, the sound speed can differ from unity and vary with time, as is the case for k-essence models (see also Carturan & Finelli 2002). Detecting these effects is an independent way of showing that dark energy does not consist of a cosmological constant.

Figures 7 and 8 illustrate the multipole moments in models in which $w(z)$ is identical but the sound speed varies. See DeDeo *et al.* (2003) for details. When the sound speed is near zero, there can be significant effects even when w is close to -1. That is, the equation of state is nearly equivalent to a cosmological constant, but the sound speed results in a detectable difference. The effects on the acoustics peaks and higher multipole moments are also relevant. If the quintessence density is at least 1% of the critical density at the surface of last scattering (as is the case for many tracker and k-essence models, for example), the modifications of the heights and shapes of the acoustic peaks in models where the sound speed is near zero are small but distinguishable from the effects due to variations of other cosmic parameters, as shown by Erickson *et al.* (2002). In addition, the sound speed can produce oscillations and other effects in the mass power spectrum, as shown in figure 9.

The data obtained from the MAP and Planck satellites and from redshift surveys may ultimately reveal these subtle effects. The precise behaviour of w and the sound speed is, by itself, of limited interest. But, what is extraordinarily important about the detection of any deviation from $c_s^2 = 1$ is that it would be a direct sign that the dark energy is a complex, dynamical fluid rather than an inert cosmological constant. Hence, these difficult targets are well worth pursuing.

5. Dark energy past and future

Most cosmologists have treated dark energy as a simple modification of the standard Big Bang/inflationary picture. But, maybe its discovery signals the need to re-evaluate our overall understanding of cosmic evolution.

Today, the consensus model of our cosmic history is based on the Big Bang picture combined with inflationary cosmology. This model has been subjected to an extraordinary battery of cosmological tests in the past decade, ranging from measurements of the CMB to detailed surveys of large-scale structure. The original picture, based on the Einstein–de Sitter model (a flat universe with matter density equal to the critical density), failed many of the tests, but replacement of 70% of the dark matter with a gravitationally self-repulsive dark energy produced a new consensus model in exquisite agreement with all cosmological tests. Hence, many cosmologists are prepared to declare our cosmic history a settled issue.

However, a second look suggests some cause for concern. The new consensus model now requires two periods of accelerated expansion: one in the early Universe during which the Universe doubles in size every 10^{-35} s, and now a second, in which the doubling time is 50 orders of magnitude greater. Each period of acceleration requires its own energy source, which must be finely tuned to satisfy observational constraints. The first acceleration has a well-defined purpose, to homogenize and flatten the Universe. The second was not predicted by either the Big Bang or inflationary pictures and it plays no known role in the Universe. (Of course, now that we know the matter density is less than the critical density, we need dark energy to bring the total to the critical value predicted by inflation. However, the expectation had originally been that the matter density would itself equal the critical density and that there is no dark energy.)

The recent proposal of a 'cyclic' Universe presents a whole new outlook on cosmic evolution in which dark energy plays a central role (Steinhardt & Turok 2002*a*, *b*). In this model, the conventional cosmic history is turned topsy-turvy. The Big Bang is not the beginning of time. Rather, it is a bridge to a pre-existing contracting era. The Universe undergoes a sequence of cycles in which it contracts in a big crunch and re-emerges in an expanding Big Bang, with trillions of years of evolution in between. The 'Big Bang' is moderated. The temperature and density of the Universe do not become infinite at any point in the cycle; indeed, they never exceed a finite bound (about a trillion trillion degrees). No inflation has taken place since the (last) bang. The current homogeneity and flatness were created by events that occurred before the most recent Big Bang, and the seeds for galaxy formation were created by instabilities arising as the Universe was collapsing towards a big crunch, prior to our Big Bang.

In this picture, dark energy is moved to centre stage as a key part of the engine that drives the periodic evolution of the Universe. Dark energy recurs as the dominant form of energy every cycle roughly 15 billion years after each bang, and it replaces two of the key roles of inflation. Although it causes the Universe to accelerate at a pace 100 orders of magnitude slower than inflation, by maintaining the acceleration for a trillion years or so, the dark energy homogenizes and flattens the Universe. In particular, it is the dark energy of a cycle ago that made the Universe homogeneous and flat prior to our own Big Bang.

A second critical feature of the dark energy is that it is not stable. In the cyclic picture, the dark energy naturally decays with time as the Universe expands. After perhaps a trillion years, the acceleration stops and the Universe begins to decelerate until the expansion halts altogether and a period of contraction begins. During contraction, quantum fluctuations in the contraction rate result in a nearly scale-invariant spectrum of perturbations that accounts for the temperature fluctuations of the CMB and large-scale structure.

Finally, the dark energy is responsible for ensuring that the cyclic evolution is an attractor solution to the evolution equations. If a random fluctuation were to kick the Universe away from the ideal cyclic evolution, the period of dark-energy domination 'redshifts' away the transient behaviour and drives the Universe back towards the regular cyclic solution.

To complete the picture, we should note that the cyclic model is motivated by recent developments in string theory, especially the ideas of branes and extra dimensions. In this picture, our three-dimensional Universe may be a hypersurface embedded in a space with one or more extra dimensions. In a version of string theory known as M-theory, for example, this hypersurface (a membrane-like surface known as a 'brane') constitutes one of the boundaries of the extra dimension, and another brane lies at the other boundary. The cyclic model proposes that the two branes interact with one another through gravity and the exchange of virtual strings and branes, resulting in a weak force that causes the branes to be drawn together and collide at regular integrals. Each collision causes the branes to bounce back to their original positions and creates matter and radiation whose gravitation causes the branes to begin to stretch. This represents the bang and the subsequent expansion and cooling. The expansion continues at a decelerating rate until, after 15 billion years, the matter and radiation density are so thinly spread that they become negligible compared with the potential energy of the interbrane force. This potential, then, is the dark energy that drives the period of accelerated expansion that has recently been observed. The branes stretch at an accelerating rate, thinning the matter density to a near vacuum and flattening any curvature and warps in the branes. Eventually, the weak force draws the brane together, reducing the dark energy and naturally ending the accelerated expansion. The 'contraction' that ensues is the contraction of the extra dimension. Our three dimensions (the branes) remain stretched out and the temperature and density remain nearly zero until the branes collide. And the cycle continues.

6. A final thought

The discovery of the retrograde motion of Mars was a surprise that was originally explained by introducing a minor modification of the heliocentric model. Only after Copernicus, Kepler and Newton was it recognized as the first hint of a great scientific revolution.

Perhaps there is a lesson here. The observation of cosmic acceleration has forced us to revise the Big Bang/inflationary picture. Should we believe, as most cosmologists suggest, that this is the last missing piece of the puzzle and our understanding of the Universe is virtually complete? Or have we just uncovered a deep dark secret that will revolutionize our whole view of the Universe and our place in it? I must confess to my own prejudice that the latter seems more likely.

I thank the Organizing Committee and The Royal Society for offering me the opportunity to participate in the Discussion Meeting. I also express my appreciation to the many students and colleagues who have guided my thinking and with whom I have had the pleasure of collaborating in the works cited here, including R. Brustein, R. Caldwell, R. Dave, S. DeDeo, J. Erickson, G. Huey, J. Khoury, I. Maor, J. McMahon, J. Ostriker, B. Ovrut, N. Turok and L. Wang. This work was supported in part by US Department of Energy grant DE-FG02-91ER40671.

References

Armendariz-Picon, C., Mukhanov, V. & Steinhardt, P. J. 2000 Dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration. *Phys. Rev. Lett.* **85**, 4438–4441.

Armendariz-Picon, C., Mukhanov, V. & Steinhardt, P. J. 2001 Essentials of k-essence. *Phys. Rev.* D **63**, 103510.

Bahcall, N. A. & Fan, X. 1998 The most massive distant clusters: determining Ω and σ_8. *Astrophys. J.* **504**, 1–6.

Bahcall, N. A., Lubin, L. M. & Dorman, V. 1995 Where is the dark matter? *Astrophys. J. Lett.* **447**, L81–L85.

Caldwell, R. R., Dave, R. & Steinhardt, P. J. 1998 Cosmological imprint of an energy component with general equation of state. *Phys. Rev. Lett.* **80**, 1582–1585.

Carlberg, R. G. *et al*. 1996 Galaxy cluster virial masses and Ω. *Astrophys. J.* **462**, 32–49.

Carroll, S. 1998 Quintessence and the rest of the world: suppressing long-range interactions. *Phys. Rev. Lett.* **81**, 3067–3070.

Carturan, D. & Finelli, F. 2002 Cosmological effects of a class of fluid dark energy models. Preprint astro-ph/0211626.

Dave, R., Caldwell, R. R. & Steinhardt, P. J. 2002 Sensitivity of the cosmic microwave background anisotropy to initial conditions in quintessence cosmology. *Phys. Rev.* D **66**, 023516.

DeDeo, S., Caldwell, R., Ostriker, J. & Steinhardt, P. J. 2003 Effects of the sound speed of quintessence on the microwave background and large scale structure. *Phys. Rev.* D **67**, 103509.

Dicke, R. H. & Peebles, P. J. E. 1979 *General relativity: an Einstein centenary survey* (ed. S. W. Hawking & W. Israel). Cambridge University Press.

Einstein, A. 1917 *Sitzungsber. Preuss. Akad. Wiss. Berlin* 142–152.

Erickson, J., Steinhardt, P. J., Caldwell, R. R., Mukhanov, V. F. & Armendariz-Picon, C. 2002 Measuring the speed of sound of quintessence. *Phys. Rev. Lett.* **88**, 121 301.

Huey, G., Wang, L., Dave, R., Caldwell, R. R. & Steinhardt, P. J. 1999 Resolving the cosmological missing energy problem. *Phys. Rev.* D **59**, 063005.

Krauss, L. & Turner, M. S. 1995 The cosmological constant is back. *Gen. Relat. Grav.* **27**, 1137–1144.

Maor, I., Brustein, R. & Steinhardt, P. J. 2001 Limitations in using luminosity distance to determine the equation of state of the Universe. *Phys. Rev. Lett.* **86**, 6–9.

Maor, I., Brustein, R., McMahon, J. & Steinhardt, P. J. 2002 Measuring the equation of state of the Universe: pitfalls and prospects. *Phys. Rev.* D **65**, 123003.

Ostriker, J. P. & Steinhardt, P. J. 1995 The observational case for a low-density Universe with a non-zero cosmological constant. *Nature* **377**, 600–603.

Perlmutter, S. *et al*. 1998 Cosmology from type Ia supernovae. *Bull. Am. Astr. Soc.* **29**, 1351.

Riess, A. G. *et al*. 1998 Observational evidence from supernovae for an accelerating Universe and a cosmological constant. *Astronom. J.* **116**, 1009–1038.

Steinhardt, P. J. 1997 Cosmological challenges for the 21st century. In *Critical problems in physics* (ed. V. L. Fitch & D. R. Marlow). Princeton University Press.

Steinhardt, P. J. & Turok, N. 2002*a* A cyclic model of the Universe. *Science* **296**, 1436–1439.

Steinhardt, P. J. & Turok, N. 2002*b* Cosmic evolution in a cyclic Universe. *Phys. Rev.* D **65**, 126003.

Steinhardt, P. J., Wang, L. & Zlatev, I. 1999 Cosmological tracking solutions. *Phys. Rev.* D **59**, 123504.

Weinberg, S. 2000 The cosmological constant problems. (Available at http://xxx.lanl.gov/abs/astro-ph/0005265.)

Zlatev, I., Wang, L. & Steinhardt, P. J. 1998 Quintessence, cosmic coincidence and the cosmological constant. *Phys. Rev. Lett.* **82**, 895.