CONSTANTS AND VARIATIONS: FROM ALPHA TO OMEGA

JOHN D. BARROW
DAMTP, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge CB3 0WA, UK

Abstract. We review some of the history and properties of theories for the variation of the gravitation and fine structure 'constants'. We highlight some general features of the cosmological models that exist in these theories with reference to recent quasar data that is consistent with time-variation in alpha since a redshift of 3.5. The behaviour of a simple class of varying-alpha cosmologies is outlined in the light of all the observational constraints. We discuss the key role played by non-zero vacuum energy and curvature in turning off the variation of constants in these theories and the issue of comparing extra-galactic and local observational data. We also show why black hole thermodynamics does not enable us to distinguish between time variations of different constants.

1. Introduction

There are a number of reasons why the possibility of varying constants should be taken seriously (Barrow, 2002). First, we know that the best candidates for unification of the forces of nature in a quantum gravitational environment only seem to exist in finite form if there are many more dimensions of space than the three that we are familiar with. This means that the true constants of nature are defined in higher dimensions and the three-dimensional shadows we observe are not fundamental and do not need to be constant. Any slow change in the scale of the extra dimensions would be revealed by measurable changes in our three-dimensional 'constants'. Second, we appreciate that some apparent constant might be determined partially or completely by some spontaneous symmetry-breaking processes in the very early universe. This introduces an irreducible random element into the values of those constants. They may be different in different parts of the universe. The most dramatic manifestation of this process is provided by the chaotic and eternal inflationary universe scenarios. Third, any outcome of a theory of quantum gravity will be intrinsically probabilistic. It is often imagined that the probability distributions for observables will be very sharply peaked but this may not be the case for all possibilities. Thus, the value of G or $\dot{G}$ might be predicted to be spatially varying random variables. Fourth, the non-uniqueness of the vacuum state for the universe would allow other deals of the constants to have occurred in different places. At present we have no idea why any of the constants of Nature take the numerical values they do. Fifth, the observational limits on possible variations are often very weak (although they can be made to sound strong by judicious parametrisations). For example, the cosmological limits on varying G

Astrophysics and Space Science **283:** 645–660, 2003.

tell us only that $\dot{G}/G \leq 10^{-2}H_0$, where H_0 is the present Hubble rate. However, the last reason to consider varying constants is currently the most compelling. For the first time there is a body of detailed astronomical evidence for the time variation of a traditional constant. The observational programme of Webb et al. (Murphy et al., 2001; Webb et al., 1999) has completed detailed analyses of three separate quasar absorption line data sets taken at Keck and finds persistent evidence consistent with the fine structure constant, α, having been *smaller* in the past, at $z = 1 - 3.5$. The shift in the value of α for all the data sets is given provisionally by $\Delta\alpha/\alpha = (-0.57 \pm 0.10) \times 10^{-5}$. This result is currently the subject of detailed analysis and reanalysis by the observers in order to search for possible systematic biases in the astrophysical environment or in the laboratory determinations of the spectral lines.

The first investigations of time-varying constants were those made by Lord Kelvin and others interested in possible time-variation of the speed of light at the end of the nineteenth century. In 1935 Milne devised a theory of gravity, of a form that we would now term 'bimetric', in which there were two times – one (t) for atomic phenomena, one (τ) for gravitational phenomena – linked by $\tau = \log(t/t_0)$. Milne (Milne, 1935) required that the 'mass of the universe' (what we would now call the mass inside the particle horizon $M \approx c^3G^{-1}t$) be constant. This required $G \propto t$. Interestingly, in 1937 the biologist J.B.S. Haldane took a strong interest in this theory and wrote several papers (Haldane, 1937) exploring its consequences for the evolution of life. He argued that biochemical activation energies might appear constant on the t timescale yet increase on the τ timescale, giving rise to a non-uniformity in the evolutionary process. Also at this time there was widespread familiarity with the mysterious 'large numbers' $O(10^{40})$ and $O(10^{80})$ through the work of Eddington (although they had first been noticed by Weyl (1919) – see Barrow and Tipler (1986) and Barrow (2002) for the history). These two ingredients were merged by Dirac in 1937 in a famous development (supposedly written on his honeymoon) that proposed that these large numbers $O(10^{40})$ were actually equal, up to small dimensionless factors. Thus, if we form $N \sim c^3t/Gm_n \sim 10^{80}$, the number of nucleons in the visible universe, and equate it to the square of $N_1 \sim e^2/Gm_n^2 \sim 10^{40}$, the ratio of the electrostatic and gravitational forces between two protons then we are led to conclude that one of the constants, e, G, c, h, m_n must vary with time. Dirac (1937) chose $G \propto t^{-1}$ to carry the time variation. Unfortunately, this hypothesis did not survive very long. Edward Teller (1948) pointed out that such a steep increase in G to the past led to huge increases in the Earth's surface temperature in the past. The luminosity of the Sun varies as $L \propto G^7$ and the radius of the Earth's orbit as $R \propto G^{-1}$ so the Earth's surface temperature $T_\oplus$varies as $(L/R^2)^{1/4} \propto G^{9/4} \propto t^{-9/4}$ and would exceed the boiling point of water in the pre-Cambrian era. Life would be eliminated. Gamow subsequently suggested that the time variation needed to reconcile the large number coincidences be carried by e rather than G, but again this strong variation was soon shown to be in conflict with geophysical and radioactive decay

data. This chapter was brought to an end by Dicke (1957) who pointed out that the $N \sim N_1^2$ large number coincidence was just the statement that t, the present age of the universe when our observations are being made, is of order the main sequence stellar lifetime, $t_{ms} \sim (Gm_n^2/hc)^{-1}h/m_nc^2 \sim 10^{10}yrs$, and therefore inevitable for observers made from elements heavier than hydrogen and helium. Dirac never accepted this anthropic explanation for the large number coincidences but curiously can be found making exactly the same type of anthropic argument to defend his own varying-G theory by highly improbable arguments (that the Sun accretes material periodically during its orbit of the galaxy and this extra material exactly cancels out the effects of overheating in the past) in correspondence with Gamow in 1967 (see Barrow, 2002, for fuller details).

Dirac's proposal acted as a stimulus to theorists, like Jordan, Brans and Dicke (Brans and Dicke, 1961), to develop rigorous theories which included the time variation of G self-consistently by modelling it as arising from the space-time variation of some scalar field $\phi(\mathbf{x}, t)$ whose motion both conserved energy and momentum and created its own gravitational field variations. In this respect the geometric structure of Einstein's equations provides a highly constrained environment to introduce variations of 'constants'. Whereas in Newtonian gravity we are at liberty to introduce a time-varying $G(t)$ into the law of gravity by writing

$$F = -\frac{G(t)Mm}{r^2}. \tag{1}$$

This creates a non-conservative dynamical system but can be solved fairly straightforwardly (Barrow, 1996). However, this strategy of simply 'writing in' the variation of G by merely replacing G by $G(t)$ in the equations that hold when G is a constant fails in general relativity. If we were to imagine the Einstein equations just generalise to (G_{ab} is the Einstein tensor)

$$G_{ab} = \frac{8\pi G(t)}{c^4}T_{ab}, \tag{2}$$

then taking a covariant divergence and using $\nabla^a G_{ab} = 0$, together with energy-momentum conservation ($\nabla^a T_{ab} = 0$) requires that $\nabla^a G \equiv 0$ and no variations are possible in Equation (2). Brans-Dicke theory is a familiar example of how the addition of an extra piece to T_{ab} together with the dynamics of a $G(\phi)$ fields makes a varying G theory possible. Despite the simplicity of this lesson in the context of a varying G theory the lesson was not taken on board when considering the variations of other non-gravitational constants and the literature is full of limits on their possible variation which have been derived by considering a theory in which the time-variation is just written into the equations which hold when the constant does not vary. Recently, the interest in the possibility that α varies in time has led to the first extensive exploration of simple self-consistent theories in which a variations occur through the variation of some scalar field.

2. Brans-Dicke Theories

2.1. Equations and solutions

Consider the paradigmatic case of Brans-Dicke (BD) theory (Brans and Dicke, 1961) to fix theoretical ideas about varying G. The form of the general solutions to the Friedmann metric in BD theories are fully understood (Barrow, 1997; Gurevich et al., 1973). There are three essential field equations for the evolution of BD scalar field $\phi(t)$ and the expansion scale factor $a(t)$ in a BD universe

$$3\frac{\dot{a}^2}{a^2} = \frac{8\pi\rho}{\phi} - 3\frac{\dot{a}}{a}\frac{\dot{\phi}}{\phi} + \frac{\omega_{BD}}{2}\frac{\dot{\phi}^2}{\phi^2} - \frac{k}{a^2} \tag{3}$$

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = \frac{8\pi}{3+2\omega_{BD}}(\rho - 3p) \tag{4}$$

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{5}$$

Here, ω_{BD} is the BD constant parameter and the theory reduces to general relativity in the limit $\omega_{BD} \to \infty$ and $\phi = G^{-1} \to$ constant. A general feature of the BD field equations is that any solution of general relativity for which the energy momentum tensor of matter has vanishing trace (e.g. vacuum, black body radiation, Yang-Mills, or magnetic field) is a particular ($\phi =$ constant) solution of BD theory.

The general solutions begin at high density dominated by the BD scalar field $\phi \sim G^{-1}$ and approximated are well approximated by the spatially flat vacuum ($\rho = p = 0$) solutions:

$$a(t) = t^{1/(\lambda+1)} \tag{6}$$

$$\phi(t) = \phi_0 t^{\lambda/(\lambda+1)} \tag{7}$$

$$\lambda = \frac{1+\sqrt{1+2\omega_{BD}/3}}{\omega_{BD}} \tag{8}$$

This vacuum solution is the $t \to 0$ attractor for the perfect-fluid solutions. The general perfect-fluid solutions with equation of state $p = \Gamma\rho$ and $k = 0$ can all be found. At early times they approach the vacuum solutions but at late time they approach particular power-law exact solutions (Nariai, 1969):

$$a(t) = t^{[2+2\omega_{BD}(1-\Gamma)]/[4+3\omega_{BD}(1-\Gamma^2)]} \tag{9}$$

$$\phi(t) = \phi_0 t^{[2(1-3\Gamma)/[4+3\omega(1-\Gamma^2)]} \tag{10}$$

At late times the asymptotic cosmological evolution is 'Machian' in the sense that the cosmological evolution is driven by the matter content rather than by the kinetic energy of the free ϕ field. In the radiation era this particular solution is the standard general relativity solution:

$$a(t) = t^{1/2}; \qquad \phi^{-1} \propto G = constant \tag{11}$$

For $p = 0$ the solutions have the form

$$a(t) = t^{(2-n)/3}; \qquad \phi^{-1} \propto G \propto t^{-n}, \tag{12}$$

which continues until the curvature term takes over the expansion. Here, n is related to the constant Brans-Dicke ω_{BD} parameter by

$$n \equiv \frac{2}{4 + 3\omega_{BD}} \tag{13}$$

and the usual general relativistic Einstein-de Sitter universe is obtained as $\omega_{BD} \to \infty$ and $n \to 0$. In a curvature-dominated era of expansion, as $t \to \infty$, the solutions for $\Gamma > -1/3$ approach the general relativity Milne vacuum solution with

$$a(t) = t \tag{14}$$

$$\phi \propto G^{-1} = constant \tag{15}$$

Notice how the curvature domination turns off the variation of G. All previous studies of varying G in cosmology have focussed on the $k = 0$ models and have not noticed the important role that would be played by negative curvature. The existence of negative curvature can leave little residue of G variation in the universe today. It also highlights the usefulness of having constraints from different cosmic times and redshifts.

The other important lesson to learn from the cosmological limits on varying G is that care must be taken when using local limits on ω_{BD}, say from light-bending by the Sun or the other solar system tests of general relativity, and then assuming that they can be used in cosmological models. In reality the evolution of the universe is inhomogeneous and there are very large variations in density between the solar system and the extragalactic universe. If we had a perfect numerical simulation of cosmology with a varying G we would be able to determine the contours of G and $\dot{G}$ with position in the universe. Until we have more information of that sort from models it is unwise to assume that the rate of change of G in the solar system will be the same as it is on cosmological scales.

An interesting particular example of this problem is given by the power-law solutions above for the case with $\Gamma = -1$. This is equivalent to the universe being dominated by a vacuum energy and leads to power-law accelerated expansion in BD theory with

$$a(t) = t^{\frac{1}{2}+\omega} \tag{16}$$

$$G \propto \phi^{-1} \propto t^{-2} \tag{17}$$

Thus it appears that if our universe were to be expanding today (as observations of the recession of type I supernovae indicate (Perlmutter et al., 1999)) then G must be falling very rapidly (even faster than Dirac predicted) locally. Clearly this is impossible observationally (Scharre and Will). The flaw in the argument is that

the $p = -\rho$ stress does not dominate the dynamics of the solar system and we must not apply the cosmological solution for the variation of G in the solar system anymore than we should apply the cosmological solution for the variation of ρ to the solar system.

It would be very interesting to find realistic solutions (exact, approximate, or numerical) for inhomogeneous cosmological models with $\phi(\mathbf{x}, t)$ in order to obtain some perspective on the likely variation in the change in G from solar system to galactic and extragalactic scales. At present, potentially the strongest cosmological limit on time-varying G is stronger than solar system tests and comes, somewhat surprisingly, from the power spectrum of galaxy clustering. The effect of varying G is to shift the cosmic epoch of equality between the matter and radiation densities which determines the location of the peak of the clustering power spectrum (Liddle et al., 1998).

3. A Simple Varying-Alpha Theory

We are going to consider some of the cosmological consequences of a simple theory with time varying α. Such a theory was first formulated by Bekenstein (Bekenstein, 1982) as a generalisation of Maxwell's equations but ignoring the consequences for the gravitational field equations. Recently, Magueijo, Sandvik and myself have completed this theory (Barrow et al., 2002a,b,c; Magueijo et al., 2002) to include the coupling to the gravitational sector and analysed its general cosmological consequences. This theory considers only a variation of the electromagnetic coupling and so far ignores any unification with the strong and electroweak interactions. We shall not discuss simultaneous variation of the electromagnetic and gravitational constants although that analysis can be done and is presented elsewhere (during the dust era of a flat Friedmann universe with varying $\alpha(t)$ and $G(t)$, their time-evolution approaches an attractor in which the product αG is a constant and $\alpha \propto G^{-1} \propto t^n$, where n is given by Equation (13).)

Our aim in studying this theory is to build up understanding of the effects of the expansion on varying α and to identify features that might carry over into more general theories in which all the unified interactions vary (Banks et al.; Langacker et al.; Marciano, 1984). The constraint imposed on varying α by the need to bring about unification at high energy is likely to be significant but the complexities of analysing the simultaneous variation of all the constants involved in the supersymmetric version of the standard model are considerable. At the most basic level we recognise that any time variation in the fine structure could be carried by either or both of the electromagnetic or weak couplings above the electroweak scale.

The idea that the charge on the electron, or the fine structure constant, might vary in cosmological time was proposed in 1948 by Teller (1948), who suggested that $\alpha \propto (\ln t)^{-1}$ was implied by Dirac's proposal that $G \propto t^{-1}$ and the numerical coincidence that $\alpha^{-1} \sim \ln(hc/Gm_{pr}^2)$, where m_{pr} is the proton mass. Later, in

1967, Gamow (1967) suggested $\alpha \propto t$ as an alternative to Dirac's time-variation of the gravitation constant, G, as a solution of the large numbers coincidences problem and in 1963 Stanyukovich had also considered varying α, (Stanyukovich, 1963), in this context. However, this power-law variation in the recent geological past was soon ruled out by other evidence (Dyson, 1967).

There are a number of possible theories allowing for the variation of the fine structure constant, α. In the simplest cases one takes c and $\hbar$ to be constants and attributes variations in α to changes in e or the permittivity of free space (see Barrow and Magueijo (1999) for a discussion of the meaning of this choice). This is done by letting e take on the value of a real scalar field which varies in space and time (for more complicated cases, resorting to complex fields undergoing spontaneous symmetry breaking, see the case of fast tracks discussed in Magueijo (2000)). Thus $e_0 \rightarrow e = e_0\epsilon(x^\mu)$, where ϵ is a dimensionless scalar field and e_0 is a constant denoting the present value of e. This operation implies that some well established assumptions must give way (Landau et al., 2001). Nevertheless, the principles of local gauge invariance and causality are maintained, as is the scale invariance of the ϵ field (under a suitable choice of dynamics). In addition there is no conflict with local Lorentz invariance or covariance. Note that $\alpha = \exp(2\psi)$

With this set up in mind, the dynamics of our theory is then constructed as follows. Since e is the electromagnetic coupling, the ϵ field couples to the gauge field as ϵA_μ in the Lagrangian and the gauge transformation which leaves the action invariant is $\epsilon A_\mu \rightarrow \epsilon A_\mu + \chi_{,\mu}$, rather than the usual $A_\mu \rightarrow A_\mu + \chi_{,\mu}$. The gauge-invariant electromagnetic field tensor is therefore

$$F_{\mu\nu} = \frac{1}{\epsilon}\left((\epsilon A_\nu)_{,\mu} - (\epsilon A_\mu)_{,\nu}\right), \tag{18}$$

which reduces to the usual form when ϵ is constant. The electromagnetic part of the action is still

$$S_{em} = -\int d^4x\sqrt{-g}F^{\mu\nu}F_{\mu\nu}. \tag{19}$$

and the dynamics of the ϵ field are controlled by the kinetic term

$$S_\epsilon = -\frac{1}{2}\frac{\hbar}{l^2}\int d^4x\sqrt{-g}\frac{\epsilon_{,\mu}\epsilon^{,\mu}}{\epsilon^2}, \tag{20}$$

as in dilaton theories. Here, l is the characteristic length scale of the theory, introduced for dimensional reasons. This constant length scale gives the scale down to which the electric field around a point charge is accurately Coulombic. The corresponding energy scale, $\hbar c/l$, has to lie between a few tens of MeV and Planck scale, $\sim 10^{19}GeV$ to avoid conflict with experiment.

Our generalisation of the scalar theory proposed by Bekenstein (1982) described in Barrow et al. (2002a,b,c); Magueijo et al. (2002) includes the gravitational effects of ψ and gives the field equations:

$$G_{\mu\nu} = 8\pi G\left(T^{matter}_{\mu\nu} + T^{\psi}_{\mu\nu} + T^{em}_{\mu\nu}e^{-2\psi}\right). \tag{21}$$

The stress tensor of the ψ field is derived from the lagrangian $\mathcal{L}_\psi = -\frac{\omega}{2}\partial_\mu\psi\partial^\mu\psi$ and the ψ field obeys the equation of motion

$$\Box\psi = \frac{2}{\omega}e^{-2\psi}\mathcal{L}_{em} \tag{22}$$

where we have defined the coupling constant $\omega = (c)/l^2$. This constant is of order ~ 1 if, as in Sandvik et al. (2002), the energy scale is similar to Planck scale. It is clear that $\mathcal{L}_{em}$ vanishes for a sea of pure radiation since then $\mathcal{L}_{em} = (E^2 - B^2)/2 = 0$. We therefore expect the variation in α to be driven by electrostatic and magnetostatic energy-components rather than electromagnetic radiation.

In order to make quantitative predictions we need to know how much of the non-relativistic matter contributes to the RHS of Equation (22). This is parametrised by $\zeta \equiv \mathcal{L}_{em}/\rho$, where ρ is the energy density, and for baryonic matter $\mathcal{L}_{em} = E^2/2$. For protons and neutrons ζ_p and ζ_n can be *estimated* from the electromagnetic corrections to the nucleon mass, 0.63 MeV and -0.13 MeV, respectively (Dvali and Zaldarriaga). This correction contains the $E^2/2$ contribution (always positive), but also terms of the form $j_\mu a^\mu$ (where j_μ is the quarks' current) and so cannot be used directly. Hence we take a guiding value $\zeta_p \approx \zeta_n \sim 10^{-4}$. Furthermore the cosmological value of ζ (denoted ζ_m) has to be weighted by the fraction of matter that is non-baryonic, a point ignored in the literature (Bekenstein, 1982). Hence, ζ_m depends strongly on the nature of the dark matter and can take both positive and negative values depending on which of Coulomb-energy or magnetostatic energy dominates the dark matter of the Universe. It could be that $\zeta_{CDM} \approx -1$ (superconducting cosmic strings, for which $\mathcal{L}_{em} \approx -B^2/2$), or $\zeta_{CDM} \ll 1$ (neutrinos). BBN predicts an approximate value for the baryon density of $\Omega_B \approx 0.03$ with a Hubble parameter of $h_0 \approx 0.6$, implying $\Omega_{CDM} \approx 0.3$. Thus depending on the nature of the dark matter ζ_m can be virtually anything between -1 and $+1$. The uncertainties in the underlying quark physics and especially the constituents of the dark matter make it difficult to impose more certain bounds on ζ_m.

We should not confuse this theory with other similar variations. Bekenstein's theory does not take into account the stress energy tensor of the dielectric field in Einstein's equations, and their application to cosmology. Dilaton theories predict a global coupling between the scalar and all other matter fields. As a result they predict variations in other constants of nature, and also a different dynamics to all the matter coupled to electromagnetism. An interesting application of our approach has also recently been made to braneworld cosmology in (Youm, 2002).

3.1. THE COSMOLOGICAL EQUATIONS

Assuming a homogeneous and isotropic Friedmann metric with expansion scale factor $a(t)$ and curvature parameter k in Equation (21), we obtain the field equa-

tions ($c \equiv 1$)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left(\rho_m\left(1+\zeta_m \exp\left[-2\psi\right]\right) + \rho_r \exp\left[-2\psi\right] + \frac{\omega}{2}\dot{\psi}^2\right) \tag{23}$$

$$-\frac{k}{a^2} + \frac{\Lambda}{3}, \tag{24}$$

where Λ is the cosmological constant. For the scalar field we have the propagation equation,

$$\ddot{\psi} + 3H\dot{\psi} = -\frac{2}{\omega}\exp\left[-2\psi\right]\zeta_m \rho_m, \tag{25}$$

where $H \equiv \dot{a}/a$ is the Hubble expansion rate. We can rewrite this more simply as

$$(\dot{\psi}a^3)^{\cdot} = N\exp[-2\psi] \tag{26}$$

where N is a positive constant defined by

$$N = -\frac{2\zeta_m \rho_m a^3}{\omega} \tag{27}$$

Note that the sign of the evolution of ψ is dependent on the sign of ζ_m. Since the observational data is consistent with a *smaller* value of α in the past, we will in this paper confine our study to *negative* values of ζ_m, in line with our recent discussion in Barrow et al. (2002a,b,c); Magueijo et al. (2002). The conservation equations for the non-interacting radiation and matter densities are

$$\dot{\rho_m} + 3H\rho_m = 0 \tag{28}$$

$$\dot{\rho}_r + 4H\rho_r = 2\dot{\psi}\rho_r. \tag{29}$$

and so $\rho_m \propto a^{-3}$ and $\rho_r\, e^{-2\psi} \propto a^{-4}$, respectively. If additional non-interacting perfect fluids satisfying equation of state $p = (\gamma - 1)\rho$ are added to the universe then they contribute density terms $\rho \propto a^{-3\gamma}$ to the RHS of Equation (24) as usual. This theory enables the cosmological consequences of varying e, to be analysed self-consistently rather than by changing the constant value of e in the standard theory to another constant value, as in the original proposals made in response to the large numbers coincidences.

We have been unable to solve these equations in general except for a few special cases (Barrow and Mota). However, as with the Friedmann equation of general relativity, it is possible to determine the overall pattern of cosmological evolution in the presence of matter, radiation, curvature, and positive cosmological constant by matched approximations. We shall consider the form of the solutions to these equations when the universe is successively dominated by the kinetic energy of the scalar field ψ, pressure-free matter, radiation, negative spatial curvature, and positive cosmological constant. Our analytic expressions are checked by numerical solutions of (24) and (25).

3.2. OBSERVATIONAL IMPLICATIONS

There are a number of conclusions that can be drawn from the study of the simple BSBM models with $\zeta_m < 0$. These models give a good fit to the varying α implied by the QSO data of Murphy et al. (2001) and Webb et al. (1999). There is just a single parameter to fit and this is given by the choice

$$-\frac{\zeta_m}{\omega} = (2 \pm 1) \times 10^{-4} \tag{30}$$

The simple solutions predict a slow (logarithmic) time increase in α during the dust era of $k = 0$ Friedmann universes. The cosmological constant turns off the time-variation of α at the redshift when the universe begins to accelerate ($z \sim 0.3$) and so there is no conflict between the α variation seen in quasars at $z \sim 1 - 3.5$ and the limits on possible variation of α deduced from the operation of the Oklo natural reactor (Shylakhter, 1976) (even assuming that the cosmological variation applies unchanged to the terrestrial environment). The reactor operated 1.8 billion years ago at a redshift of only $z \sim 0.1$ when no significant variations were occurring in α. The slow logarithmic increase in α also means that we would not expect to have seen any effect yet in the anisotropy of the microwave backgrounds (Battye et al.; Avelino 2000): the value of α at the last scattering redshift, $z = 1000$, is only 0.005% lower than its value today. Similarly, the essentially constant evolution of α predicted during the radiation era leads us to expect no measurable effects on the products of Big Bang nucleosynthesis (BBN) (Barrow, 1987) because α was only 0.007% smaller at BBN than it is today. This does not rule out the possibility that unification effects in a more general theory might require variations in weak and strong couplings, or their contributions to the neutron-proton mass difference, which might produce observable differences in the light element productions and new constraints on varying α at $z \sim 10^9 - 10^{10}$. By contrast varying alpha cosmologies with $\zeta > 0$ lead to bad consequences . The fine structure falls rapidly at late times and the variation is such that it even comes to dominate the Friedmann equation for the cosmological dynamics. We regard this as a signal that such models are astrophysically ruled out and perhaps also mathematically badly behaved.

We should also mention that theories in which α varies will in general lead to violations of the weak equivalence principle (WEP). This is because the α variation is carried by a field like ψ and this couples differently to different nuclei because they contain different numbers of electrically charged particles (protons). The theory discussed here has the interesting consequence of leading to a relative acceleration of order 10^{-13} (Magueijo et al.) if the free coupling parameter is fixed to the value given in Equation (30) using a best fit of the theories cosmological model to the QSO observations of Murphy et al. (2001); Webb et al. (1999). Other predictions of such WEP violations have also been made in Damour and Polyakov (1994); Dvali and Zaldarriaga (a, 2002); Damour et al. (2002). The observational upper bound on this parameter is just an order of magnitude larger, at 10^{-12}, but space-based tests

planned for the STEP mission are expected to achieve a sensitivity of order 10^{-18} and will provide a completely independent check on theories of time-varying e and α.This is an exciting prospect for the future.

3.3. THE NATURE OF THE FRIEDMANN SOLUTIONS

The cosmological behaviour of the solutions to these equations was studied by us in detail, both analytically and numerically (Barrow et al., 2002a,b,c; Magueijo et al., 2002; Barrow and Mota). Typically, the variation in α does not have a significant effect on the evolution of the scale factor at late times although the cosmological expansion does significantly affect the evolution of α. The evolution of α is summarised as follows:

During the radiation era $a(t) \sim t^{1/2}$ and α is constant in universes with our entropy per baryon and present value of α like our own. It increases in the dust era, where $a(t) \sim t^{2/3}$. The increase in α however, is very slow with a late-time solution for ψ proportional to $\frac{1}{2}\log(2N\log(t))$, and so

$$\alpha \sim 2N \log t \tag{31}$$

This slow increase continues until the expansion becomes dominated by negative curvature, $a(t) \sim t$, or by a cosmological vacuum energy, $a(t) \sim \exp[\Lambda t/3]$. Thereafter α asymptotes rapidly to a constant. If we set the cosmological constant equal to zero and $k = 0$ then, during the dust era, α would continue to increase indefinitely. The effect of the expansion is very significant at all times. If we were to turn it off and set $a(t)$ constant then we could solve the ψ equation to give the following exponentially growing evolution for α, (Barrow and Mota):

$$\alpha = \exp[2\psi] = A^{-2}\cosh^2[AN^{1/2}(t + t_0)];\ A\ constant. \tag{32}$$

From these results it is evident that non-zero curvature or cosmological constant brings to an end the increase in the value of α that occurs during the dust-dominated era. Hence, if the spatial curvature and Λ are both too *small* it is possible for the fine structure constant to grow too large for biologically important atoms and nuclei to exist in the universe. There will be a time in the future when α reaches too large a value for life to emerge or persist. The closer a universe is to flatness, or the closer Λ is to zero, so the longer the monotonic increase in α will continue, and the more likely it becomes that life will be extinguished. Conversely, a non-zero positive Λ or a non-zero negative curvature will stop the increase of α earlier and allow life to persist for longer. If life can survive into the curvature or Λ-dominated phases of the universe's history then it will not be threatened by the steady cosmological increase in α unless the universe collapses back to high density.

This type of behaviour can also be found in the presence of time-varying G. If a BD dust universe is exactly flat ($k = 0$) then G will continue to fall forever. Only if there is negative curvature will the evolution of G eventually be turned off and the expansion asymptote to the Milne behaviour with $a = t$ and $G \rightarrow$ constant. Again,

without the small deviation from flatness the strength of gravity would ultimately become too weak for the existence of stars and planets and the universe would become biologically inhospitable, if not uninhabitable.

There have been several studies, following Carter (1974) and Tryon (1973), of the need for life-supporting universes to expand close to the 'flat' Einstein de Sitter trajectory for long periods of time. This ensures that the universe cannot collapse back to high density before galaxies, stars, and biochemical elements can form by gravitational instability, or expand too fast for stars and galaxies to form by gravitational instability (Collins and Hawking, 1973; Barrow and Tipler, 1986). Likewise, it was pointed out by Barrow and Tipler (1986) that there are similar anthropic restrictions on the magnitude of any cosmological constant, Λ. If it is too large in magnitude it will either precipitate premature collapse back to high density (if $\Lambda < 0$) or prevent the gravitational condensation of any stars and galaxies (if $\Lambda > 0$). Thus existing studies provide anthropic reasons why we can expect to live in an old universe that is neither too far from flatness nor dominated by a much stronger cosmological constant than observed ($|\Lambda| \leq 10\,|\Lambda_{obs}|$).

Inflationary universe models provide a possible theoretical explanation for proximity to flatness but no explanation for the smallness of the cosmological constant. Varying speed of light theories (Moffat, 1993; Albrecht and Magueijo, 1999; Barrow, 1999; Barrow and Magueijo, 1999; Magueijo, 2001) offer possible explanations for proximity to flatness and smallness of a classical cosmological constant (but not necessarily for one induced by vacuum corrections in the early universe). We have shown that if we enlarge our cosmological theory to accommodate variations in some traditional constants then *it appears to be anthropically disadvantageous for a universe to lie too close to flatness or for the cosmological constant to be too close to zero*. This conclusion arises because of the coupling between time-variations in constants like α and the curvature or Λ, which control the expansion of the universe. The onset of a period of Λ or curvature domination has the property of dynamically stabilising the constants, thereby creating favourable conditions for the emergence of structures. This point has been missed in previous studies because they have never combined the issues of Λ and flatness and the issue of the values of constants. By coupling these two types of anthropic considerations we find that too small a value of Λ or the spatial curvature can be as poisonous for life as too much. Universes like those described above, with increasing $\alpha(t)$, lead inexorably to an epoch where α is too large for the existence of atoms, molecules, and stars to be possible.

Surprisingly, there has been almost no consideration of habitability in cosmologies with time-varying constants since Haldane's discussions (Haldane, 1937) of the biological consequences of Milne's bimetric theory of gravity. Since then, attention has focussed upon the consequences of universes in which the constants are different but still constant. Those cosmologies with varying constants that have been studied have not considered the effects of curvature or Λ domination on the variation of constants and have generally considered power-law variation to hold

for all times. The examples described here show that this restriction has prevented a full appreciation of the coupling between the expansion dynamics of the universe and the values of the constants that define the course of local physical processes within it. Our discussion of a theory with varying α shows for the first time a possible reason why the 3-curvature of universes and the value of any cosmological constant may need to be bounded *below* in order that the universe permit atomic life to exist for a significant period. Previous anthropic arguments have shown that the spatial curvature of the universe and the value of the cosmological constant must be bounded *above* in order for life-supporting environments (stars) to develop. We note that the lower bounds discussed here are more fundamental than these upper bounds because they derive from changes in α which have direct consequences for biochemistry whereas the upper bounds just constrain the formation of astrophysical environments by gravitational instability. Taken together, these arguments suggest that within an ensemble of all possible worlds where α and G are time variables, there might only be a finite interval of *non-zero* values of the curvature and cosmological constant contributions to the dynamics that both allow galaxies and stars to form and their biochemical products to persist.

3.4. THE ROLE OF INHOMOGENEITIES

We can also detect where and how we might expect spatial variations to arise in a fuller description. Aside from the complexities of the full inhomogeneous cosmological solution for the formation of galaxies, stars, and planets, we can isolate non-uniformities that enter through the constant parameter N which dictates the form and time-evolution of $a(t)$ and $\alpha(t)$. First we see that N is proportional to the density of electromagnetically charged matter in the universe. This will possess some spatial variation and is of order 10^{-5} on large scales. More significant though is the variation of the baryonic content of the CDM density with scale. We need the CDM to be dominated by matter with magnetic charge (but see Bekenstein). This can be the case on large scale but we know that the dark matter becomes dominated by baryons (therefore with $\zeta > 0$) locally. Hence, there is expected to be a very significant spatial variation of ζ with scale, including a change of sign, which will feed into the variation of α.

3.5. GENERAL PROPERTIES OF THE EVOLUTION OF ALPHA AND G

The evolution equation for $\psi(t)$ has a number of simple but important properties. Since $N > 0$ the right-hand side of Equation (26) must be positive. This means that linearisations of this equation are dangerous and give rise to linearisation instabilities unless attention is confined to the regime $\psi << 1$. In general the positivity property means that there can be no oscillations of ψ or α in time in solutions of this equation. This follows from the required positivity of $(\dot{\psi} a^3\dot{)}$, which means that ψcannot have a maximum. The possible cosmological evolutions for ψ and α are decrease to a minimum followed by a monotonic increase, monotonic decrease, or

monotonic increase. This conclusion holds independently of the value of k in the Friedmann equation. This has one very important consequence. It means that the asymptotic monotonic non-decrease of α found in our flat and open universes will still occur in closed universes. There cannot be a sudden change in the evolution of α when the universe starts to collapse. This also means that if we model spherical overdensities by closed universes embedded in a flat background then the evolution of $\alpha(t)$ in the overdensities will be very similar to that in the flat background even when the overdensities collapse to form bound 'clusters'. This has the important implication that such an inhomogeneous universe will not end up with very different values of α and $\dot{\alpha}$ in inside and outside the bound inhomogeneities.

This argument can also be applied to the evolution of ϕ and G in BD theory. Consider the case of dust ($p = 0$). The combination $(\dot{\phi}a^3)^{\cdot}$ must now be positive and so ϕ cannot have a maximum and G cannot have a minimum regardless of the sign of the curvature parameter k. In particular, $G(t)$ cannot oscillate. Again, this property acts as a safeguard on the divergent evolution of G inside and outside overdensities.

4. The Second Law

There has been considerable recent discussion (Youm, 2002; M. Duff) about the equivalence of models of the variation of different dimensional 'constants' of Nature. In particular, it has been suggested that consideration of the second law of black hole thermodynamics distinguishes, say, variations of e from variations of c and that some of these variations could be ruled out because they bring about a decrease in time of the Bekenstein-Hawking entropy of a charged black hole. Others have argued that no such distinction is operationally possible. However, we believe that the most crucial factor has been missed in this discussion. In theories which generalise general relativity by allowing traditional constants (like G or e) to vary the black hole solutions with event horizons are particular solutions of the theory in which the constant concerned is a constant. When the constant varies the black hole solution no longer exists and there is no longer any black hole thermodynamics to constrain the variation. The situation is very clear in the simple case of a Schwarzschild black hole in Brans Dicke theory. We know from the work of Hawking (1972) that the black hole solutions are the same as those in general relativity. Thus Schwarzschild is a $\phi \propto G^{-1} =$ constant solution of the Brans-Dicke field equations. The entropy of this black hole is

$$S_{bh} \propto GM^2. \tag{33}$$

If we were to apply the second law to this formula it would appear to say that all cosmological solutions in which G falls with time are ruled out. However, this would not be a correct deduction (which is fortunate because we see from Equation (10) that essentially all Brans-Dicke cosmologies have such behaviour) because

ϕ and G are *constant* on the Schwarzschild horizon. If we allow variation of G then the solution turns into a naked singularity and the thermodynamic relations no longer exist. Thus one cannot at present use considerations of black hole thermodynamics to constrain or distinguish the time or space variation of constants of Nature by simply 'writing in' time variations into the formulae that define the black hole when these constants do not vary.

Acknowledgements

I would like to thank my collaborators João Magueijo, Håvard Sandvik, John Webb, Michael Murphy and David Mota for their essential contributions to the work described here. I would also like to thank Bruce Bassett, Thibault Damour, Paul Davies, Tamara Davies, Thomas Dent, Carlos Martins, John Moffat and Clifford Will for discussions.

References

Albrecht, A. and Magueijo, J.: 1999, *Phys. Rev. D* **59**, 043516; Barrow, J.D. and Magueijo, J.: 1998, *Phys. Lett. B* **443**, 104.
Avelino, P.P. et al.: 2001, *Phys. Rev. D* **62**, 123508, and 2001: astro-ph/0102144.
Banks, T., Dine, M. and Douglas, M.R.: hep-ph/0112059.
Barrow, J.D. and Mota, D.: gr-qc/0207012.
Barrow, J.D.: 1999, *Phys. Rev. D* **59**, 043515.
Barrow, J.D.: 1987, *Phys. Rev. D* **35**, 1805.
Barrow, J.D.: 2002, *The Constants of Nature: from Alpha to Omega,* Jonathan Cape, London.
Barrow, J.D.: 1996, *MNRAS* **282**, 1397.
Barrow, J.D.: 1997, in: N. Sanchez (ed.), Proc. Erice Summer School *Current Topics in Astrofundamental Physics: Primordial Cosmology*, pp. 269–305, gr-qc/9711084.
Barrow, J.D. and Magueijo, J.: 1999, *Phys. Lett. B* **447**, 246.
Barrow, J.D. and Tipler, F.J.: 1986, *The Anthropic Cosmological Principle*, Oxford UP, Oxford.
Barrow, J.D., Sandvik, H.B. and Magueijo, J.: 2002a, *Phys. Rev. D* **66**, 043515.
Barrow, J.D., Sandvik, H.B. and Magueijo, J.: 2002b, *Phys. Rev. D* **65**, 123501.
Barrow, J.D., Sandvik, H.B. and Magueijo, J.: 2002c, *Phys. Rev. D* **65**, 063504.
Battye, R., Crittenden, R. and Weller, J.: *Phys. Rev. D* **63**, 043505.
Bekenstein, J.D.: 1982, *Phys. Rev. D* **25**, 1527.
Brans, C. and Dicke, R.H.: 1961, *Phys. Rev.* **124**, 924.
Carter, B.: 1974, in: M.S. Longair (ed.), *Confrontation of Cosmological Theories with Observation*, Reidel, Dordrecht, p. 291 and *Large Numbers in Astrophysics and Cosmology*, unpublished preprint, Inst. Theoretical Astronomy, 1968.
Collins, C.B. and Hawking, S.W.: 1973, *Astrophys. J.* **180**, 317.
Damour, T., Piazza, F. and Veneziano, G.: 2002, *Phys. Rev. D 66*, 046007 and *Phys. Rev. Lett.* **89**, 081601.
Damour, T. and Polyakov, A.: 1994, *Nucl. Phys. B* **423**, 532.
Dicke, R.H.: 1957, *Rev. Mod. Phys.* **29**, 355 and 1961: *Nature* **192**, 440.
Dirac, P.A.M.: 1937, *Nature* **139**, 323.
Duff, M.: hep-th/0208093.

Dvali, G. and Zaldarriaga, M.: a, hep-ph/0108217.
Dvali, G.R. and Zaldarriaga, M.: 2002, *Phys. Rev. Lett.* **88**, 091303.
Dyson, F.: 1967, *Phys. Rev. Lett.* **19**, 1291.
Gamow, G.: 1967, *Phys. Rev. Lett.* **19**, 759.
Gurevich, L., Finkelstein, A.M. and Ruban, V.A.: 1973, *Astrophys. Space Sci.* **22**, 231.
Haldane, J.B.S.: 1937, *Nature* **139**, 1002 and 1944: **158**, 555 and article in 1955: M.L. Johnson et al. (eds.), *New Biology* **16**, Penguin, London.
Hawking, S.W.: 1972, *Comm. Math. Phys.* **25**, 152.
Landau, S., Sisterna, P. and Vucetich, H.: 2001, *Phys. Rev. D* **63**, 081303(R).
Langacker, P., Segre, G. and Strassler, M.: hep-ph/0112233; Calmet, X. and Fritzsch, H.: hep-ph/0112110; Fairbairn, H. and Dent, T.: hep-ph/0112279.
Liddle, A.R., Barrow, J.D. and Mazumdar, A.: 1998, *Phys. Rev. D* **58**, 027302.
Magueijo, J.: 2001, *Phys. Rev. D* **63**, 043502.
Magueijo, J., Barrow, J.D. and Sandvik, H.: astro-ph/0202374.
Magueijo, J.: 2000, *Phys. Rev. D* **62**, 103521.
Magueijo, J., Barrow, J.D. and Sandvik, H.B.: 2002, *Phys. Lett. B* **541**, 201.
Marciano, W.: 1984, *Phys. Rev. Lett.* **52**, 489; Banks, T., Dine, M. and Douglas, M.R.: 2002, *Phys. Rev. Lett.* **88** 131301; Langacker, P., Segre, G. and Strassler, M.: 2002, *Phys. Lett. B* **528** 121–128; Calmet, X. and Fritzsch, H.: hep-ph/0112110.
Milne, E.A.: 1935, *Relativity, Gravitation and World Structure,* Clarendon, Oxford.
Moffat, J.: 1993, *Int. J. Mod. Phys. D* **2**, 351 and astro-ph/0109350.
Murphy, M., Webb, J., Flambaum, V., Dzuba, V., Churchill, C., Prochaska, J. and Wolfe, A.: 2001, *MNRAS* **327**, 1208.
Nariai, H.: 1969: *Prog. Theo. Phys.* **42**, 544.
Perlmutter, S. et al.: 1999, *Ap. J.* **517**, 565; Perlmutter, S. et al.: 1997, *Ap. J.* **483**, 565; Perlmutter, S. et al.: 1998, *Nature* **391**, 51; Garnavich, P.M. et al.: 1998, *Ap. J. Lett.* **493**, L53; Schmidt, B.P.: 1998, *Ap. J.* **507**, 46; Riess, A.G. et al.: 1998, *AJ* **116**, 1009.
Sandvik, H.B., Barrow, J.D. and Magueijo, J.: 2002, *Phys. Rev. Lett.* **88**, 031302.
Scharre, P.D. and Will, C.: gr-qc/0109044 and Will, C.: 1993, *Theory and Experiment in Gravitational Physics*, CUP, Cambridge.
Shylakhter, A.: 1976, *Nature* **264**, 340; also Fujii, Y. et al.: 2000, *Nucl. Phys. B* **573**, 377; and Damour, T. and Dyson, F.J.: 1996, *Nucl. Phys. B* **480**, 37.
Stanyukovich, K.P.: 1963, *Sov. Phys. Dokl.* **7**, 1150.
Teller, E.: 1948, *Phys. Rev.* **73**, 801.
Tryon, E.: 1973, *Nature* **246**, 387.
Webb, J.K., Flambaum, V.V., Churchill, C.W., Drinkwater, M.J. and Barrow, J.D.: 1999, *Phys. Rev. Lett.* **82**, 884; Webb, J.K., Murphy, M.T., Flambaum, V.V., Dzuba, V.A., Barrow, J.D., Churchill, C.W., Prochaska, J.X. and Wolfe, A.M.: 2001, *Phys. Rev. Lett.* **87**, 091301.
Weyl, H.: 1919, *Ann. Physik* **59**, 129.
Youm, D.: 2002, *Phys. Rev. D* **66**, 043506; also Davies, P.C.W., Davies, T. and Lineweaver, C.: 2002, *Nature* **418**, 602.
Youm, D.: 2002, *Mod. Phys. Lett. A* **17**, 175.