Carnegie Observatories Astrophysics Series, Vol. 2:
Measuring and Modeling the Universe, 2004
ed. W. L. Freedman (Cambridge: Cambridge Univ. Press)

Cosmology and Life

MARIO LIVIO
Space Telescope Science Institute

Abstract

I examine some recent findings in cosmology and their potential implications for the emergence of life in the Universe. In particular, I discuss the requirements for carbon-based life, anthropic considerations with respect to the nature of dark energy, the possibility of time-varying constants of nature, and the question of the rarity of intelligent life.

1.1 Introduction

The progress in cosmology in the past few decades leads also to new insights into the global question of the emergence of intelligent life in the Universe. Here I am not referring to discoveries that are related to very localized regions, such as the detection of extrasolar planetary systems, but rather to properties of the Universe at large.

In order to set the stage properly for the topics to follow, I would like to start with four observations with which essentially all astronomers agree. These four observations *define* the cosmological context of our Universe.

(1) Ever since the observations of Vesto Slipher in 1912–1922 (Slipher 1917) and Hubble (1929), we know that the spectra of distant galaxies are redshifted.
(2) Observations with the *Cosmic Background Explorer (COBE)* have shown that, to a precision of better than 10^{-4}, the cosmic microwave background (CMB) is *thermal*, at a temperature of 2.73 K (Mather et al. 1994).
(3) Light elements, such as deuterium and helium, have been synthesized in a high-temperature phase in the past (e.g., Gamow 1946; Alpher, Bethe, & Gamow 1948; Hoyle & Tayler 1964; Peebles 1966; Wagoner, Fowler, & Hoyle 1967).
(4) Deep observations, such as the Hubble Deep Field, have shown that galaxies in the distant Universe look younger. Namely, their sizes are smaller (e.g., Roche et al. 1996), and there is a higher fraction of irregular morphologies (e.g., Abraham et al. 1996). This is what one would expect from a higher rate of interactions, and from "building blocks" of today's galaxies.

When the above four observational facts are combined and considered together, there is no escape from the conclusion that our Universe is *expanding and cooling*. This conclusion is entirely *consistent* with the "hot big bang" model. Sometimes the stronger statement, that these observations "prove" that there was a hot big bang, is made. However, the scientific method does not truly produce "proofs."

During the past decade, deep observations with a variety of ground-based and space-based observatories have advanced our understanding of the history of the Universe far beyond the mere statement that a big bang had occurred (see, e.g., the determination of cosmological parameters by the *Wilkinson Microwave Anisotropy Probe (WMAP)*; Spergel et al. 2003). In particular, remarkable progress has been achieved in the understanding of the cosmic star formation history.

Using different observational tracers (e.g., the UV luminosity density) of star formation in high-redshift galaxies, tentative plots for the star formation rate (SFR) as a function of redshift have been produced (e.g., Lilly et al. 1996; Madau et al. 1996; Steidel et al. 1999). There is little doubt that the SFR rises from the present to about $z \approx 1$. What happens in the redshift range $z \approx 1$–5 is still somewhat controversial. While some studies suggest that the SFR reaches a peak at $z \approx 1$–2 and then declines slightly toward higher redshifts (e.g., Steidel et al. 1999; or maybe even more than slightly, Stanway, Bunker, & McMahon 2003), or stays fairly flat up to $z \approx 5$ (e.g., Calzetti & Heckman 1999; Pei, Fall, & Hauser 1999), others claim that the SFR continues to rise to $z \approx 8$ (Lanzetta et al. 2002). The latter claim is based on the suggestion that previous studies had failed to account for surface brightness dimming effects. For my present purposes, however, it is sufficient that the history of the *global* SFR is on the verge of being determined (if it has not been determined already). A knowledge of the SFR as a function of redshift allows for the first time for meaningful constraints to be placed on the global emergence of carbon-based life.

1.2 Remarks about Carbon-based Life

The main contributors of carbon to the interstellar medium are intermediate-mass (1–8 $M_\odot$) stars (e.g., Wood 1981; Yungelson, Tutukov, & Livio 1993; Timmes, Woosley, & Weaver 1995), through the asymptotic giant branch and planetary nebulae phases. A knowledge of the cosmic SFR history, together with a knowledge of the initial mass function (presently still uncertain for high redshift), therefore allows for an approximate calculation of the rate of carbon production as a function of redshift (Livio 1999). For a peaked SFR, of the type obtained by Madau et al. (1996), for example, the peak in the carbon production rate is somewhat delayed (by $\lesssim 1$ billion years) with respect to the SFR peak. The decline in the carbon production rate is also shallower for $z \lesssim 1$ (than the decline in the SFR), owing to the buildup of a stellar reservoir in the earlier epochs.

Assuming a "principle of mediocrity," one would expect the emergence of most carbon-based life in the Universe to be perhaps not too far from the peak in the carbon production rate—around $z \approx 1$ (for a peak in the SFR at $z \approx 1$–2). Since the time scale required to develop intelligent civilizations may be within a factor of 2 of the lifetime of F5 to mid-K stars (the ones possessing continuously habitable zones; Kasting, Whitmore, & Reynolds 1993; and see § 1.5 below), it can be expected that intelligent civilizations have emerged when the Universe was $\gtrsim 10$ Gyr old. A younger emergence age may be obtained if the SFR does not decline at redshifts $1.2 \lesssim z \lesssim 8$ (e.g., Lanzetta et al. 2002).

Carbon features in most anthropic arguments. In particular, it is often argued that the existence of an excited state of the carbon nucleus (the 0_2^+ state) is a manifestation of fine-tuning of the constants of nature, which allowed for the appearance of carbon-based life.

Carbon is formed through the triple-α process in two steps. In the first, two α particles form the unstable (lifetime $\sim 10^{-16}$ s) ^{8}Be. In the second, a third α particle is captured, via $^8\text{Be}(\alpha,\gamma)^{12}\text{C}$. Hoyle argued that in order for the 3α reaction to proceed at a rate sufficient to produce the observed cosmic carbon, a resonant level must exist in ^{12}C, a few hundred keV above the ^{8}Be + ^{4}He threshold. Such a level was indeed found experimentally (Dunbar et al. 1953; Hoyle, Dunbar, & Wenzel 1953; Cook, Fowler, & Lauritsen 1957).

The question of how fine-tuned this level needs to be for the existence of carbon-based life has been the subject of considerable research. The most recent work on this topic was done by Oberhummer and collaborators (e.g., Oberhummer, Csótó, & Schlattl 2000; Csótó, Oberhummer, & Schlattl 2001; Schlattl et al. 2004). These authors used a model that treats the ^{12}C nucleus as a system of 12 interacting nucleons, with the approximate resonant reaction rate

$$r_{3\alpha} = 3^{3/2} N_\alpha^3 \left(\frac{2\pi\hbar^2}{M_\alpha k_B T} \right)^3 \frac{\Gamma_\gamma}{\hbar} \exp\left(-\frac{\varepsilon}{k_B T} \right) . \tag{1.1}$$

Here M_α and N_α are the mass and number density of α particles, respectively, ε is the resonance energy (in the center-of-mass frame), Γ_γ is the relative width, and all other symbols have their usual meaning. Oberhummer et al. introduced small variations in the strengths of the nucleon-nucleon interaction and in the fine structure constant (affecting ε and Γ_γ), and calculated stellar models using the modified rates. In their initial work, Oberhummer et al. (2000) concluded that a change of more than 0.5% in the strength of the strong interaction or more than 4% in the strength of the electromagnetic interaction would result in essentially no production of carbon or oxygen [considering the ^{12}C$(\alpha,\gamma)^{16}$O and ^{16}O$(\alpha,\gamma)^{20}$Ne reactions] in any star. More specifically, a decrease in the strong-interaction strength by 0.5%, coupled with an increase in the fine structure constant by 4%, resulted in a decrease in the carbon production by a factor of a few tens in 20 $M_\odot$ stars, and by a factor of $\sim$100 in 1.3 $M_\odot$ stars. Taken at face value, this seemed to support anthropic claims for extreme fine-tuning necessary for the emergence of carbon-based life.

Earlier calculations by Livio et al. (1989) indicated less impressive fine-tuning. Livio et al. showed that shifting (artificially) the energy of the carbon resonant state by up to 0.06 MeV does not result in a significant reduction in the production of carbon. Since this 0.06 MeV should be compared to the *difference* between the resonance energy in ^{12}C and the 3α threshold (calculated with the basic nucleon-nucleon interaction), it was not obvious that a particularly fantastic fine-tuning was required. Most recently, however, Schlattl et al. (2004) reinvestigated the dependence of carbon and oxygen production in stars on the 3α rate. These authors found that following the entire stellar evolution was crucial. They concluded that in massive stars the C and O production strongly depends on the initial mass. In intermediate- and low-mass stars, Schlattl et al. found that the high carbon production during He shell flashes leads to a *lower* sensitivity of the C and O production on the 3α rate than inferred by Oberhummer et al. (2000). Schlattl et al. (2004) concluded by saying that "fine-tuning with respect to the obtained carbon and oxygen abundance is more complicated and far less spectacular" than that found by Oberhummer et al. (2000).

1.3 Dark Energy and Life

In 1998, two teams of astronomers, working independently, presented evidence that the expansion of the Universe is accelerating (Riess et al. 1998; Perlmutter et al. 1999). The evidence was based primarily on the unexpected faintness (by $\sim$0.25 mag) of distant ($z \approx 0.5$) Type Ia supernovae, compared to their expected brightness in a universe decelerating under its own gravity. The results favored values of $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$ for the matter and "dark energy" density parameters, respectively. Subsequent observations of the supernova SN 1997ff, at the redshift of $z \simeq 1.7$, strengthened the conclusion of an accelerating Universe (Riess et al. 2001). This supernova appeared *brighter* relative to SNe in a coasting universe, as expected from the fact that at $z \approx 1.7$ a universe with $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$ would still be in its decelerating phase. The observations of SN 1997ff do not support any alternative interpretation (such as dust extinction or evolutionary effects) in which supernovae are expected to dim monotonically with redshift. Measurements of the power spectrum of the cosmic microwave background (e.g., Abroe et al. 2002; de Bernardis et al. 2002; and Netterfield et al. 2002; and, most recently the *WMAP* results, Bennett et al. 2003) provide strong evidence for flatness ($\Omega_m + \Omega_\Lambda = 1$). When combined with estimates of Ω_m based on mass-to-light ratios, X-ray temperatures of intracluster gas, and dynamics of clusters (all of which give $\Omega_m \lesssim 0.3$; e.g., Strauss & Willick 1995; Carlberg et al. 1996; Bahcall et al. 2000), again a value of $\Omega_\Lambda \approx 0.7$ is obtained.

Arguably the two greatest puzzles physics is facing today are:

(1) Why is the dark energy (vacuum energy) density, ρ_v, so small, but not zero? (Or, why does the vacuum energy gravitate so little?)
(2) Why *now*? Namely, why do we find at present that $\Omega_\Lambda \approx \Omega_m$?

The first question reflects the fact that taking graviton energies up to the Planck scale, M_P, would produce a dark energy density

$$\rho_v \approx M_P^4 \approx (10^{18}\ \mathrm{GeV})^4, \tag{1.2}$$

which misses the observed one, $\rho_v \approx (10^{-3}\ \mathrm{eV})^4$, by more than 120 orders of magnitude. Even if the energy density in fluctuations in the gravitational field is taken only up to the supersymmetry-breaking scale, M_{SUSY}, we still miss the mark by a factor of 60 orders of magnitude, since $\rho_v \approx M_{\mathrm{SUSY}}^4 \approx (1\ \mathrm{TeV})^4$. Interestingly, though, a scale $M_v \approx (M_{\mathrm{SUSY}}/M_P)M_{\mathrm{SUSY}}$ produces the right order of magnitude. However, while a few attempts in this direction have been made (e.g., Arkani-Hamed et al. 2000), no satisfactory model has been developed.

The second question is related to the anti-Copernican fact that Ω_Λ may be associated with a cosmological constant, while Ω_m declines continuously (and in any case, ρ_v may be expected to have a different time behavior from ρ_m), and yet the first time that we are able to measure both reliably, we find that they are of the same order.

The attempts to solve these problems fall into three general categories:

(1) The behavior of "quintessence" fields
(2) Alternative theories of gravity
(3) Anthropic considerations

The attempts of the first type have concentrated in particular on "tracker" solutions (e.g., Zlatev, Wang, & Steinhardt 1998; Albrecht & Skordis 2002), in which the smallness of Ω_Λ is a direct consequence of the Universe's old age. Generally, a uniform scalar field, ϕ, is taken to evolve according to

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \tag{1.3}$$

where $V'(\phi) = \frac{dV}{d\phi}$ and H is the Hubble parameter. The energy density of the scalar field is given by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \tag{1.4}$$

and that of matter and radiation, ρ_m, by

$$\dot{\rho}_m = -3H(\rho_m + P_m), \tag{1.5}$$

where P_m is the pressure. For a potential of the form

$$V(\phi) = \phi^{-\alpha} M^{4+\alpha}, \tag{1.6}$$

where $\alpha > 0$ and M is an adjustable constant ($M \ll M_P$), and a field that is initially much smaller than the Planck mass, one obtains a solution in which a transition occurs from an early ρ_m-dominance to a late ρ_ϕ-dominance (with no need to fine-tune the initial conditions). Nevertheless, for the condition $\rho_\phi \approx \rho_m$ to actually be satisfied at the present time requires (Weinberg 2001) that the parameter M would satisfy

$$M^{4+\alpha} \simeq (8\pi G)^{-1-\alpha/2} H_0^2, \tag{1.7}$$

which is not easily explicable.

In order to overcome this problem, some quintessence models choose potentials in which the Universe has periodically been accelerating in the past (e.g., Dodelson, Kaplinghat, & Stewart 2000), so that the dark energy's dominance today appears naturally.

A very different approach regards the accelerating expansion not as being propelled by dark energy, but rather as being the result of a modified gravity. For example, models have been developed (Deffayet, Dvali, & Gabadadze 2002), in which ordinary particles are localized on a three-dimensional surface (3-brane) embedded in infinite-volume extra dimensions to which gravity can spread. The model is constructed in such a way that observers on the brane discover Newtonian gravity (four dimensional) at distances that are shorter than a crossover scale, r_c, which can be of astronomical size. In one version, the Friedmann equation is replaced by

$$H^2 + \frac{k}{a^2} = \left(\sqrt{\frac{\rho}{3M_P^2} + \frac{1}{4r_c^2}} + \epsilon\frac{1}{2r_c^2}\right)^2, \tag{1.8}$$

where ρ is the total energy density, a is the scale factor and $\epsilon = \pm 1$.

In this case, the dynamics of gravity are governed by whether ρ/M_P^2 is larger or smaller than $1/r_c^2$. Choosing $r_c \approx H_0^{-1}$ preserves the usual cosmological results. At large cosmic distances, however, gravity spreads into extra dimensions (the force law becomes five dimensional), and becomes weaker—directly affecting the cosmic expansion. Basically, at late times, the model has a self-accelerating cosmological branch with $H = 1/r_c$ (to leading-order Equation 1.8 can be parameterized as $H^2 - H/r_c \simeq \rho/3M_P^2$). Interestingly, it has recently been suggested that the viability of these models can be tested by lunar ranging experiments (Dvali, Gruzinov, & Zaldarriaga 2003). I should also note that the *WMAP* results indicated an intriguing lack of correlated signal on angular scales greater than 60 degrees (Spergel et al. 2003), reinforcing the low quadrupole seen already in *COBE* results. One possible, although at this stage speculative, interpretation of these results is that they signal the breakdown of conventional gravity on large scales.

A third class of proposed solutions to the dark energy problems relies on anthropic selection effects, and therefore on the *existence* of intelligent life in our Universe. The basic premise of this approach is that some of the constants of nature are actually random variables, whose range of values and *a priori* probabilities are nevertheless determined by the laws of physics. The observed big bang, in this picture, is simply one member of an ensemble. It is further assumed that a "principle of mediocrity" applies; namely, we can expect to observe the most probable values (Vilenkin 1995). Using this approach, Garriga, Livio, & Vilenkin (2000; following the original idea of Weinberg 1987) were able to show that when the cosmological constant Λ is the only variable parameter, the order of magnitude coincidence $t_0 \approx t_\Lambda \approx t_G$ (where t_0 is the present time; t_Λ is the time Ω_Λ starts to dominate; t_G is the time when giant galaxies were assembled) finds a natural explanation (see also Bludman 2000).

Qualitatively, the argument works as follows.

In a geometrically flat universe with a cosmological constant, gravitational clustering can no longer occur after redshift $(1+z_\Lambda) \approx (\rho_\Lambda/\rho_{m0})^{1/3}$ (where ρ_{m0} is the present matter density). Therefore, requiring that ρ_Λ does not dominate before redshift $z_{\rm max}$, at which the earliest galaxies formed, requires (e.g., Weinberg 1987)

$$\rho_\Lambda \lesssim (1+z_{\rm max})^3 \rho_{m0}. \tag{1.9}$$

One can expect the *a priori* (independent of observers) probability distribution $P(\rho_\Lambda)$ to vary on some characteristic scale, $\Delta\rho_\Lambda \approx \eta^4$, determined by the underlying physics. Irrespective of whether η is determined by the Planck scale ($\sim 10^{18}$ GeV), the grand unification scale ($\sim 10^{16}$ GeV) or the electroweak scale ($\sim 10^2$ GeV), $\Delta\rho_\Lambda$ exceeds the anthropically allowed range of ρ_Λ (Eq. 1.9) by so many orders of magnitude that it looks reasonable to assume that

$$P(\rho_\Lambda) = const, \tag{1.10}$$

over the range of interest. Garriga & Vilenkin (2001) and Weinberg (2001) have shown that this assumption is satisfied by a broad class of models, even though not automatically. With a flat distribution, a value of ρ_Λ picked randomly (and which may characterize a "pocket" universe) from an interval $|\rho_\Lambda| \lesssim \rho_\Lambda^{\rm max}$, will, with a high probability, be of the order of $\rho_\Lambda^{\rm max}$. The principle of mediocrity, however, means that we should observe a value of ρ_Λ that maximizes the number of galaxies. This suggests that we should observe the largest value of ρ_Λ that is still consistent with a substantial fraction of matter having collapsed into galaxies—in other words, $t_\Lambda \approx t_G$, as observed. In § 1.2 I argued that the appearance of carbon-based life may be associated roughly with the peak in the star formation rate, $t_{\rm SFR}$. The "present time," t_0, is not much different from that (in that it takes only a fraction of a stellar lifetime to develop intelligent life), hence $t_0 \approx t_{\rm SFR}$. Finally, hierarchical structure formation models suggest that vigorous star formation is closely associated with the formation of galactic-size objects (e.g., Baugh et al. 1998; Fukugita, Hogan, & Peebles 1998). Therefore, $t_G \approx t_{\rm SFR}$, and we obtain $t_0 \approx t_G \approx t_\Lambda$.

Garriga et al. (2000) further expanded their discussion to treat not just Λ, but also the density contrast at recombination, $\sigma_{\rm rec}$, as a random variable (see also Tegmark & Rees 1998). The galaxy formation in this case is spread over a much wider time interval, and proper account has to be taken for the fact that the cooling of protogalactic clouds collapsing at very late times is too slow for efficient fragmentation and star formation (fragmentation occurs if the cooling time scale is shorter than the collapse time scale, $\tau_{\rm cool} < \tau_{\rm grav}$). Assuming an *a priori* probability distribution of the form

$$P(\sigma_{\rm rec}) \sim \sigma_{\rm rec}^{-\alpha}, \tag{1.11}$$

Garriga et al. found that "mediocre" observers will detect $\sigma_{\rm rec} \approx 10^{-4}$, $t_0 \approx t_G \approx t_\Lambda \approx t_{\rm cb}$, as observed, *if* $\alpha > 3$ (here the "cooling boundary" $t_{\rm cb}$ is the time after which fragmentation is suppressed).

Other anthropic explanations for the value of the cosmological constant and the "why now?" problem have been suggested in the context of maximally extended ($N = 8$) supergravity (Kallosh & Linde 2003; Linde 2003). In particular, the former authors found that the Universe can have a sufficiently long lifetime only if the scaler field satisfies initially $|\phi| \lesssim M_P$, and if the value of the potential $V(0)$, which plays the role of the cosmological constant, does not exceed the critical density $\rho_0 \approx 10^{-120} M_P^4$.

Personally, I feel that anthropic explanations to the dark energy problems should be regarded as the *last resort*, only after all attempts to find explanations based on first principles have been exhausted and failed. Nevertheless, the anthropic explanation may prove to be the correct one, if our understanding of what is truly *fundamental* is lacking. A historical example can help to clarify this last statement. Johannes Kepler (1571–1630) was obsessed by the following two questions:

(1) Why were there precisely six planets? (only Mercury, Venus, Earth, Mars, Jupiter and Saturn were known at his time)
(2) What was it that determined that the planetary orbits would be spaced as they are?

The first thing to realize is that these "why" and "what" questions were a novelty in the astronomical vocabulary. Astronomers before Kepler were usually satisfied with simply recording the observed positions of the planets; Kepler was seeking a theoretical explanation. Kepler finally came up with preposterously fantastic (and absolutely wrong) answers to his two questions in *Mysterium Cosmographicum*, published in 1597. He suggested that the reason for there being six planets is that there are precisely five Platonic solids. Taken as boundaries (with an outer spherical boundary corresponding to the fixed stars), the solids create six spacings. By choosing a particular order for the solids to be embedded in each other, with the Earth separating the solids that can stand upright (cube, tetrahedron, and dodecahedron) from those that "float" (octahedron and icosahedron), Kepler claimed to have explained the sizes of the orbits too (the spacings agreed with observations to within 10%).

Today we recognize what was the *main* problem with Kepler's model—Kepler did not understand that neither the number of planets nor their spacings are *fundamental* quantities that need to have an explanation from first principles. Rather, both are the result of historical accidents in the solar protoplanetary disk. Still, it is perfectly legitimate to give an anthropic "explanation" for the Earth's orbital radius. If that orbit were not in the continuously habitable zone around the sun (Kasting & Reynolds 1993), we would not be here to ask the question.

It is difficult to admit it, but our current model for the composition of the Universe: $\sim$73% dark energy, $\sim$23% cold dark matter, $\sim$4% baryonic matter, and maybe $\sim$0.5% neutrinos, appears no less preposterous than Kepler's model. While some version of string (or *M*–) theories may eventually provide a first-principles explanation for all of these values, it is also possible, in my opinion, that these individual values are in fact not fundamental, but accidental. Maybe the only fundamental property is the fact that *all the energy densities add up to produce a geometrically flat universe*, as predicted by inflation (Guth 1981; Hawking

1982; Steinhardt & Turner 1984) and confirmed by *WMAP* (Spergel et al. 2003). Clearly, for any anthropic explanation of the value of Ω_Λ to be meaningful at all, even in principle, one requires the existence of a large ensemble of universes, with different values of Ω_Λ. That this requirement may actually be fulfilled is precisely one of the consequences of the concept of "eternal inflation" (Steinhardt 1983; Vilenkin 1983; Linde 1986; Goncharov, Linde, & Mukhanov 1987; Linde 2003). In most inflationary models the time scale associated with the expansion is much shorter than the decay time scale of the false vacuum phase, $\tau_{\rm exp} \ll \tau_{\rm dec}$. Consequently, the emergence of a fractal structure of "pocket universes" surrounded by false vacuum material is almost inevitable (Garcia-Bellido & Linde 1995; Guth 2001; for a different view, see, e.g., Bucher, Goldhaber, & Turok 1995; Turok 2001).

This ensemble of pocket universes may serve as the basis on which anthropic argumentation can be constructed (even though the definition of probabilities on this infinite set is nontrivial; see, e.g., Linde, Linde, & Mezhlumian 1995; Vilenkin 1998).

1.4 Varying Constants of Nature?

Another recent finding, which, *if confirmed*, may have implications for the emergence of life in the Universe, is that of cosmological evolution of the fine structure constant $\alpha \equiv e^2/\hbar c$ (Webb et al. 1999, 2001, and references therein). Needless to say, life as we know it places significant anthropic constraints on the range of values allowed for α. For example, the requirement that the lifetime of the proton would be longer than the main sequence lifetime of stars results in an upper bound $\alpha \lesssim 1/80$ (Ellis & Nanopoulos 1981; Barrow, Sandvik, & Magueijo 2002a). The claimed detection of time variability was based on shifts in the rest wavelengths of redshifted UV resonance transitions observed in quasar absorption systems. Basically, the dependence of observed wave number at redshift z, w_z, on α can be expressed as

$$w_z = w_0 + a_1 w_1 + a_2 w_2, \tag{1.12}$$

where a_1 and a_2 represent relativistic corrections for particular atomic masses and electron configurations, and

$$w_1 = \left(\frac{\alpha_z}{\alpha_0}\right)^2 - 1 \tag{1.13}$$

$$w_2 = \left(\frac{\alpha_z}{\alpha_0}\right)^4 - 1. \tag{1.14}$$

Here α_0 and α_z represent the present day and redshift z values of α, respectively. By analyzing a multitude of absorption lines from many multiplets in different ions, such as Fe II and Mg II transitions in 28 absorption systems (in the redshift range $0.5 \lesssim z \lesssim 1.8$), and Ni II, Cr II, Zn II, and Si IV transitions in some 40 absorption systems (in the redshift range $1.8 \lesssim z \lesssim 3.5$), Webb et al. (2001) concluded that α was *smaller* in the past. Their data suggest a 4σ deviation

$$\frac{\Delta\alpha}{\alpha} = -0.72 \pm 0.18 \times 10^{-5} \tag{1.15}$$

over the redshift range $0.5 \lesssim z \lesssim 3.5$ (where $\Delta\alpha/\alpha = \frac{\alpha_z - \alpha_o}{\alpha_o}$). It should be noted, however, that the data are consistent with *no* variation for $z \lesssim 1$, in agreement with many previous studies (e.g., Bahcall, Sargent, & Schmidt 1967; Wolfe, Brown, & Roberts 1976; Cowie & Songaila 1995).

Murphy et al. (2001) conducted a comprehensive search for systematic effects that could potentially be responsible for the result (e.g., laboratory wavelength errors, isotopic abundance effects, heliocentric corrections during the quasar integration, line blending, and atmospheric dispersion). While they concluded that isotopic abundance evolution and atmospheric dispersion could have an effect, this was in the direction of actually amplifying the variation in α [to $\Delta\alpha/\alpha = (-1.19 \pm 0.17) \times 10^{-5}$]. The most recent results of Webb et al. are not inconsistent with limits on α from the Oklo natural uranium fission reactor (which was active 1.8×10^9 years ago, corresponding to $z \approx 0.1$) and with constraints from experimental tests of the equivalence principle. The former suggests $\Delta\alpha/\alpha \simeq (-0.4 \pm 1.4) \times 10^{-8}$ (Fuji et al. 2000), and the latter *allows* for a variation of the magnitude observed in the context of a general dynamical theory relating variations of α to the electromagnetic fraction of the mass density in the Universe (Bekenstein 1982; Livio & Stiavelli 1998).

Before going any further, I would like to note that what is desperately needed right now is an independent confirmation (or refutation) of the results of Webb et al. by other groups, both through additional (and preferably different) observations and via independent analysis of the data. In this respect it is important to realize that the reliability of the SNe Ia results (concerning the accelerating Universe) was enormously enhanced by the fact that two separate teams (the Supernova Cosmology Project and the High-z Supernova Team) reached the same conclusion independently, using different samples and different data analysis techniques. A first small step in the direction of testing the variable α result came from measurements of the CMB. A likelihood analysis of BOOMERanG and MAXIMA data, allowing for the possibility of a time-varying α (which, in turn, affects the recombination time) found that in general the data may prefer a smaller α in the past (although the conclusion is not free of degeneracies; Avelino et al. 2000; Battye, Crittenden, & Weller 2001). A second, much more important step, came through an extensive analysis using the nebular emission lines of [O III] $\lambda\lambda$4959, 5007 Å (Bahcall, Steinhardt, & Schlegel 2004). Bahcall et al. found $\Delta\alpha/\alpha = (-2 \pm 1.2) \times 10^{-4}$ (corresponding to $|\alpha^{-1} d\alpha/dt| < 10^{-13}$ yr^{-1}, which they consider to be a null result, given the precision of their method) for quasars in the redshift range $0.16 < z < 0.8$. While this result is not formally inconsistent with the variation claimed by Webb et al., the careful analysis of Bahcall et al. has cast some serious doubts on the ability of the "many-multiplet" method employed by Webb and his collaborators to actually reach the accuracy required to measure fractional variations in α at the 10^{-5} level. For example, Bahcall et al. have shown that to achieve that precision, one needs to assume that the velocity profiles of different ions in different clouds are essentially the same to within 1 km s^{-1}. Clearly, much more work on this topic is needed. I should also note right away that, in order not to be in conflict with the yield of ^{4}He, $|\Delta\alpha/\alpha|$ cannot exceed $\sim 2 \times 10^{-2}$ at the time of nucleosynthesis (e.g., Bergström, Igury, & Rubinstein 1999).

On the theoretical side, simple cosmological models with a varying fine structure constant have now been developed (e.g., Sandvik, Barrow, & Magueijo 2003; Barrow, Sandvik, & Magueijo 2002b). They share some properties with Kaluza-Klein-type models in which α varies at the same rate as the extra dimensions of space (e.g., Damour & Polyakov 1994), and with varying-speed-of-light theories (e.g., Albrecht & Magueijo 1999; Barrow & Magueijo 2000).

The general equations describing a geometrically flat, homogeneous, isotropic, variable-α universe are (Beckenstein 1982; Livio & Stiavelli 1998; Sandvik et al. 2002) the Friedmann equation (with $G = c \equiv 1$)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\left[\rho_m\left(1+|\zeta_m|e^{-2\psi}\right)+\rho_r e^{-2\psi}+\rho_\psi+\rho_\Lambda\right], \tag{1.16}$$

the evolution of the scalar field varying α ($\alpha = \exp(2\psi)e_0^2/\hbar c$)

$$\ddot{\psi}+3H\dot{\psi} = -\frac{2}{w}e^{-2\psi}\zeta_m\rho_m, \tag{1.17}$$

and the conservation equations for matter and radiation

$$\dot{\rho}_m + 3H\rho_m = 0 \tag{1.18}$$

$$\dot{\rho}_r + 4H\rho_r = 2\dot{\psi}\rho_r. \tag{1.19}$$

Here, ρ_m, ρ_r, ρ_ψ, ρ_Λ are the densities of matter, radiation, scalar field ($\frac{w}{2}\dot{\psi}^2$), and vacuum, respectively, $a(t)$ is the scale factor ($H \equiv \dot{a}/a$), $w = \hbar c/l^2$ is the coupling constant of the dynamic Langrangian (l is a length scale of the theory), and ξ_m is a dimensionless parameter that represents the fraction of mass in Coulomb energy of an average nucleon compared to the free proton mass.

Equations 1.16–1.19 were solved numerically by Sandvik et al. (2002) and Barrow et al. (2002), assuming a negative value of the parameter ξ_m/w, and the results are interesting both from a purely cosmological point of view and from the perspective of the emergence of life. First, the results are consistent with both the claims of a varying α of Webb et al. (which, as I noted, badly need further confirmation) and with the more secure, by now, observations of an accelerating Universe (Riess et al. 1998; Perlmutter et al. 1999; Spergel et al. 2003), while complying with the geological and nucleosynthetic constraints. Second, Barrow et al. find that α remains almost constant in the radiation-dominated era, experiences a small logarithmic time increase during the matter-dominated era, but approaches a constant value again in the Λ-dominated era. This behavior has interesting anthropic consequences. The existence of a nonzero vacuum energy contribution is now *required* in this picture to dynamically stabilize the fine structure constant. In a universe with zero Λ, α would continue to grow in the matter-dominate era to values that would make the emergence of life impossible (Barrow et al. 2001).

Clearly, the viability of all of the speculative ideas above relies at this point on the confirmation or refutation of time-varying constants of nature.

1.5 Is Intelligent Life Extremely Rare?

With the discovery of $\sim$100 massive extrasolar planets (Mayor & Queloz 1995; Marcy & Butler 1996, 2000), the question of the potential existence of extraterrestrial, Galactic, intelligent life has certainly become more intriguing than ever. This topic has attracted much attention and generated many speculative (by necessity) probability estimates. Nevertheless, in a quite remarkable paper, Carter (1983) concluded on the basis of the near-equality between the lifetime of the sun, $t_\odot$, and the time scale of biological evolution on Earth, t_ℓ, that extraterrestrial intelligent civilizations are exceedingly rare in the Galaxy. Most significantly, Carter's conclusion is supposed to hold even if the conditions optimal for the emergence of life are relatively common.

Let me reproduce here, very briefly, Carter's argument. The basic, and very crucial assumption on which the argument is based is that the lifetime of a star, t_*, and the time scale of biological evolution on a planet around that star, t_ℓ (taken here, for definiteness, to be the time scale for the appearance of complex land life), are *a priori entirely independent*. In other words, the assumption is that land life appears at some *random* time with respect to the main sequence lifetime of the star. Under this assumption, one expects that generally one of the two relations $t_\ell \gg t_*$ or $t_\ell \ll t_*$ applies (the set where $t_\ell \approx t_*$ is of negligible measure for two independent quantities). Let us examine each one of these possibilities. If *generally* $t_\ell \ll t_*$, it is very difficult to understand why in the first system found to contain complex land life, the Earth-Sun system, the two time scales are nearly equal, $t_\ell \approx t_*$. If, on the other hand, *generally* $t_\ell \gg t_*$, then clearly the first system we find must exhibit $t_\ell \approx t_*$ (since for $t_\ell \gg t_*$ complex land life would not have developed). Therefore, one has to conclude that *typically* $t_\ell \gg t_*$, and that consequently, complex land life will generally not develop—the Earth is an extremely rare exception.

Carter's argument is quite powerful and not easily refutable. Its basic assumption (the independence of t_ℓ and t_*) appears on the face of it to be solid, since t_* is determined primarily by nuclear burning reactions, while t_ℓ is determined by biochemical reactions and the evolution of species. Nevertheless, the fact that the star is the main energy source for biological evolution (light energy exceeds the other sources by 2–3 orders of magnitude; e.g., Deamer 1997), already implies that the two quantities are not completely independent.

Let me first take a purely mathematical approach and examine what would it take for the condition $t_\ell \approx t_*$ to be satisfied in the Earth-Sun system *without* implying that extraterrestrial intelligent life is extremely rare. Imagine that t_ℓ and t_* are not independent, but rather that

$$t_\ell/t_* = f(t_*), \tag{1.20}$$

where $f(t_*)$ is some *monotonically increasing* function in the narrow range $t_*^{\min} \lesssim t_* \lesssim t_*^{\max}$ that allows the emergence of complex land life through the existence of continuously habitable zones (corresponding to stellar spectral types F5 to mid-K; Kasting et al. 1993). Note that, for a Salpeter (1955) initial mass function, the distribution of stellar lifetimes behaves as

$$\psi(t_*) \approx t_*. \tag{1.21}$$

Consequently, if relation 1.20 were to hold, it would in fact be the *most probable* that in the first place where we encounter an intelligent civilization we would find that $t_\ell/t_* \approx 1$, as in the Earth-Sun system. In other words, if we could identify some processes that are likely to produce a monotonically increasing $t_* - t_\ell/t_*$ relation, then the near equality of t_ℓ and t_* in the Earth-Sun system would find a natural explanation, with no implications whatsoever for the frequency of intelligent civilizations. A few years ago, I proposed a simple toy-model for how such a relation might arise (Livio 1999). The toy-model was based on the assumption that the appearance of land life has to await the build-up of a sufficient layer of protective ozone (Berkner & Marshall 1965; Hart 1978), and on the fact that oxygen in a planet's atmosphere is released in the first phase from the dissociation of water (Hart 1978; Levine, Hayes, & Walker 1979). Given that the duration of this phase is inversely proportioned to the intensity of radiation in the 1000–2000 Å range, a relation between t_ℓ and t_* can be established. In fact, a simple calculation gave

$$t_\ell/t_* \simeq 0.4(t_*/t_\odot)^{1.7}, \tag{1.22}$$

precisely the type of monotonic relation needed.

I should be the first to point out that the toy-model above is nothing more than that—a toy model. It does point out, however, that at the very least, establishing a link between the biochemical and astrophysical time scales may not be impossible. Clearly, the emergence of complex life on Earth required many factors operating together. These include processes that appear entirely accidental, such as the stabilization of the Earth's tilt against chaotic evolution by the Moon (e.g., Laskar, Joutel, & Boudin 1993). Nevertheless, we should not be so arrogant as to conclude everything from the one example we know. The discovery of many "hot Jupiters" (giant planets with orbital radii $\lesssim 0.05$ AU) has already demonstrated that the solar system may not be typical. We should keep an open mind to the possibility that biological complexity may find other paths to emerge, making various "accidents," coincidences, and fine-tuning unnecessary. In any case, the final scientific assessment on life in the Universe will probably come from biologists and observers—not from speculating theorists like myself.

Acknowledgements. This work has been supported by Grant 938-COS191 from the Templeton Foundation. I am grateful to Andrei Linde and Heinz Oberhummer for helpful comments.

References

Abraham, R. G., van den Bergh, S., Glazebrook, K., Ellis, R. S., & Santiago, B. X. 1996, ApJS, 101, 1
Abroe, M. E., et al. 2002, MNRAS, 334, 11
Albrecht, A., & Magueijo, J. 1999, Phys. Rev. D, 59, 043516
Albrecht, A., & Skordis, C. 2002, Phys. Rev. Lett., 84, 2076
Alpher, R. A., Bethe, H., & Gamow, G. 1948, Phys. Rev., 73, 803
Arkani-Hamed, N., Hall, L. J., Colda, C., & Murayama, H. 2000, Phys. Rev. Lett., 85, 4434
Avelino, P. P., Martins, C. J. A. P., Rocha, G., & Viana, P. 2000, Phys. Rev. D, 62, 123508
Bahcall, J. N., Sargent, W. L. W, & Schmidt, M. 1967, ApJ, 149, L11
Bahcall, J. N., Steinhardt, C. L., & Schlegel, D. 2004, ApJ, in press (astro-ph/0301507)
Bahcall, N. A., Cen, R., Davé, R., Ostriker, J. P., & Yu, O. 2000, ApJ, 541, 1
Barrow, J. D., & Magueijo, J. 2000, ApJ, 532, L87
Barrow, J. D., Sandvik, H. B., & Magueijo, J. 2002a, Phys. Rev. D, 65, 123501
——. 2002b, Phys. Rev. D, 65, 063504
Battye, R. A., Crittenden, R., & Weller, J. 2001, Phys. Rev. D, 63, 043505
Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998, ApJ, 498, 504
Bekenstein, J. D. 1982, Phys. Rev. D, 25, 1527
Bennett, C. L., et al. 2003, ApJS, 148, 1
Bergström, L., Iguri, S., & Rubinstein, H. 1999, Phys. Rev. D, 60, 045005
Berkner, L. V., & Marshall, K. C. 1965, J. Atmos. Sci., 22, 225
Bludman, S. 2000, Nucl. Phys. A, 663–664, 865
Bucher, M., Goldhaber, A., & Turok, N. 1995, Phys. Rev. D, 52, 3314
Calzetti, D., & Heckman, T. M. 1999, ApJ, 519, 27
Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Grabel, P., Morris, S., & Pritchet, C. J. 1996, ApJ, 462, 32
Carter, B. 1983, Phil. Trans. R. Soc. London A, 310, 347
Cook, C. W., Fowler, W. A., & Lauritsen, T. 1957, Phys. Rev., 107, 508
Cowie, L. L., & Songaila, A. 1995, ApJ, 453, 596
Csótó, A., Oberhummer, H., & Schlattl, H. 2001, Nucl. Phys. A, 688, 560
Damour, T., & Polyakov, A. M. 1994, Nucl. Phys. B, 423, 532
Deamer, D. W. 1997, Microb. and Molec. Bio. Rev., 61, 239

de Bernardis, P., et al. 2002, ApJ, 564, 559
Deffayet, C., Dvali, G., & Gabadadze, G. 2002, Phys. Rev. D, 65, 044023
Dodelson, S., Kaplinghat, M., & Stewart, E. 2000, Phys. Rev. Lett., 85, 5276
Dunbar, D. N. F., Pixley, R. E., Wenzel, W. A., & Whaling, W. 1953, Phys. Rev., 92, 649
Dvali, G., Gruzinov, A., & Zaldarriaga, M. 2003, Phys. Rev. D, 68, 024012
Ellis, J. D., & Nanopoulos, D. V. 1981, Nature, 292, 436
Fuji, V., et al. 2000, Nucl. Phys. B, 573, 377
Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518
Gamow, G. 1946, Phys. Rev., 70, 527
Garcia-Bellido, J., & Linde, A. O. 1995, Phys. Rev. D, 51, 429
Garriga, J., Livio, M., & Vilenkin, A. 2000, Phys. Rev. D., 61, 023503
Garriga, J., & Vilenkin, A. 2001, Phys. Rev. D., 64, 023517
Goncharov, A. S., Linde, A. D., & Mukhanov, V. F. 1987, Int. J. Mod. Phys. A, 2, 561
Guth, A. H. 1981, Phys. Rev. D, 23, 347
Guth, A. H. 2001, in Astrophysical Ages and Time Scales, ed. T. von Hippel, C. Simpson, & N. Mansit (San Francisco: ASP), 3
Hart, M. H. 1978, Icarus, 33, 23
Hawking, S. W. 1982, Phys. Lett. B, 115, 295
Hoyle, F., Dunbar, D. N. F., & Wenzel, W. A. 1953, Phys. Rev., 92, 1095
Hoyle, F., & Tayler, R. J. 1964, Nature, 203, 1108
Hubble, E. 1929, Proc. Nat. Acad. Sci., 15, 168
Kallosh, R., & Linde, A. 2003, Phys. Rev. D., 67, 023510
Kasting, J. F., & Reynolds, R. T. 1993, Icarus, 101, 108
Kasting, J. F., Whitmore, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108
Lanzetta, K. M., Yahata, N., Pascarelle, S., Chen, H.-W., & Fernández-Soto, A. 2002, ApJ, 570, 492
Laskar, J., Joutel, F., & Boudin, F. 1993, Nature, 361, 615
Levine, J. S., Hayes, P. B., & Walker, J. C. G. 1979, Icarus, 39, 295
Lilly, S. J., Le Fèvre, O., Hammer, F., & Crampton, D. 1996, ApJ, 460, L1
Linde, A. 2003, in Science and Ultimate Reality: From Quantum to Cosmos, ed. J. D. Barrow, P. C. W. Davies, & C. L. Harper (Cambridge: Cambridge Univ. Press), in press
Linde, A., Linde, D., & Mezhlumian, A. 1995, Phys. Lett. B, 345, 203
Linde, A. D. 1986, Mod. Phys. Lett. A, 1, 81
Livio, M. 1999, ApJ, 511, 429
Livio, M., Hollowell, D., Weiss, A., & Truran, J. W. 1989, Nature, 340, 281
Livio, M., & Stiavelli, M. 1998, ApJ, 507, L13
Madau, P., Ferguson, H. C., Dickinson, M., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388
Marcy, G. W., & Butler, R. P. 1996, ApJ, 464, L147
——. 2000, PASP, 112, 137
Mather, J. C., et al. 1994, ApJ, 420, 439
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Murphy, M. T., Webb, J. K., Flambaum, V. V., Churchill, C. W., & Prochaska, J. X. 2001, MNRAS, 327, 1223
Netterfield, C. B., et al. 2002, ApJ, 571, 604
Oberhummer, H., Csótó, A., & Schlattl, H. 2000, Science, 289, 88
Peebles, P. J. E. 1966, ApJ, 146, 542
Pei, Y., Fall, S. M., & Hauser, M. G. 1999, ApJ, 522, 604
Perlmutter, S., et al. 1999, ApJ, 517, 565
Riess, A. G., et al. 1998, AJ, 116, 1009
——. 2001, ApJ, 560, 49
Roche, N., Ratnatunga, K. U., Griffiths, R. E., Im, M., & Neuschaefer, L. W. 1996, MNRAS, 282, 1247
Salpeter, E. E. 1955, ApJ, 121, 161
Sandvik, H. B., Barrow, J. D., & Magueijo, J. 2002, Phys. Rev. Lett., 88, 031302
Schlattl, H., Heger, A., Oberhummer, H., Rauscher, T., & Csótó, A. 2004, MNRAS, in press
Slipher, V. M. 1917, Proc. Amer. Phil. Soc., 56, 403
Spergel, D. N., et al. 2003, ApJS, 148, 175
Stanway, E. R., Bunker, A. J., & McMahon, R. G. 2003, MNRAS, 342, 439
Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1

Steinhardt, P. J. 1983, in The Very Early Universe, ed. G. W. Gibbons, S. Hawking, & S. T. C. Siklos (Cambridge: Cambridge Univ. Press), 251
Steinhardt, P. J., & Turner, M. S. 1984, Phys. Rev. D, 29, 2162
Strauss, M. A., & Willick, J. A. 1995, Phys. Rep., 261, 271
Tegmark, M., & Rees, M. J. 1998, ApJ, 499, 526
Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995, ApJS, 98, 617
Turok, N. 2001, in Birth and Evolution of the Universe, ed. K. Sato & M. Kawasaki (Tokyo: Universal Academy Press), 1
Vilenkin, A. 1983, Phys. Rev. D, 27, 2848
——. 1995, Phys. Rev. Lett., 74, 846
——. 1998, Phys. Rev. Lett., 81, 5501
Wagoner, R. V., Fowler, W. A., & Hoyle, F. 1967, ApJ, 148, 3
Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J., & 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. Prochanska, J. X., & Wolfe, A. M. 2001, Phys. Rev. Lett., 87, 091301
Weinberg, S. 1987, Phys. Rev. Lett., 59, 2607
——. 2001, in Sources and Detection of Dark Matter and Energy in the Universe, ed. D. B. Cline (Berlin: Springer), 18
Wolfe, A. M., Brown, R. L., & Roberts, M. S. 1976, Phys. Rev. Lett., 37, 179
Wood, P. R. 1981, in Physical Processes in Red Giants, ed. I. Iben, Jr. & A. Renzini (Dordrecht: Reidel), 205
Yungelson, L., Tutukov, A. V., & Livio, M. 1993, ApJ, 418, 794
Zlatev, I., Wang, L., & Steinhardt, P. J. 1998, Phys. Rev. Lett., 82, 896