Time-Varying α and Particle Physics

Tom Banks,[1,2,*] Michael Dine,[2,†] and Michael Douglas[1,3,‡]
[1]Department of Physics and Astronomy, Rutgers University NHETC, Piscataway, New York
[2]Department of Physics, University of California–Santa Cruz/SCIPP, Santa Cruz, California
[3]I.H.E.S., Bures-sur-Yvette 91440, France
(Received 9 January 2002; published 15 March 2002)

We argue that models in which an observable variation of the fine structure constant is explained by motion of a cosmic scalar field are not stable under renormalization and require massive fine-tuning that cannot be explained by any known mechanism.

DOI: 10.1103/PhysRevLett.88.131301 PACS numbers: 98.80.Cq, 12.15.Lk

Recent observations of distant quasars have revived the suggestion that the fine structure constant α varies over cosmological time scales. The observations of [1] suggest a variation of $\delta\alpha/\alpha \sim 10^{-4}$ over the time period since emission of the quasar light.

Within both the contexts of effective field theory and M theory, it is natural to model the change of the fine structure constant by coupling a dynamical scalar field ϕ to the photon kinetic term in the low energy effective action.

$$\delta e^{-2} = \epsilon \frac{\phi}{M} F_{\mu\nu}^2 . \tag{1}$$

Here M is a typical scale over which ϕ varies, and ϵ is an additional parameter. For typical string moduli, M might be of order the Planck or string scale, while ϵ might be of order 1, for axionlike fields which couple at tree level or one loop. On the other hand, it can also be much smaller.

One can imagine two broad classes of scenarios to explain the putative variation of α. Usually, a slow, secular variation of ϕ is postulated. Another possibility is that there was a first order phase transition between the time at which the quasar light was emitted and the present. In this Letter we argue that any such time variation of α raises difficulties. In particular, such a large variation of α can be compatible with basic principles of quantum field theory only if there is an extraordinary degree of fine-tuning of many parameters of the underlying theory. (This conclusion has also been reached by Kaloper and Susskind [2]. Closely related ideas have been discussed by Donoghue [3], who also stresses that variation of couplings implies variation of the cosmological constant and makes estimates which are similar in spirit to some of those discussed here.)

Let us first state the basic argument and later show how this comes out of concrete calculations. It is that the vacuum energy, as computed in the standard model, or in more general low energy effective field theory, must depend on α, even taking into account our ignorance about the resolution of the cosmological constant problem. This dependence can be estimated in various ways, and combining these estimates with the criterion that the new contributions do not dominate the energy densities which appear in standard cosmology leads to fantastically tight bounds.

In generic quantum field theory, varying an arbitrary dimensionless coupling such as α will lead to a variation in the vacuum energy V controlled by the cutoff scale Λ,

$$\delta V = c\,\delta\alpha\,\Lambda^4 \tag{2}$$

with c an $\mathcal{O}(1)$ constant. For example, in QED this is $\langle F_{\mu\nu}F^{\mu\nu}\rangle$ and in perturbation theory $c = 1/\pi + \ldots$. Perturbative corrections and the effects of other matter in the standard model will modify this result, but again lead to $c \sim \mathcal{O}(1)$.

The estimate (2) might well be an overestimate. If throughout the relevant cosmic history the field ϕ has been near its minimum, the vacuum energy will be of order $\delta\alpha^2$,

$$\delta V = C\alpha\left(\frac{\delta\alpha}{\alpha}\right)^2 \Lambda^4. \tag{3}$$

In models of quintessence or in which the change in α is due to a phase transition, the first order estimate is likely to be correct.

Let us be extremely conservative and take Λ to be the QCD scale $\sim$100 MeV. Then

$$\delta V = \left(\frac{\delta\alpha}{\alpha}\right) 10^{29}\ \mathrm{eV}^4. \tag{4}$$

However, the matter dominated era of conventional cosmology has a maximum energy density of order 10^4 eV^4. Thus, some of what is supposed to be the matter dominated era, and in particular the period when the quasar light was emitted, was instead dominated by a large scalar field potential energy. This changes classical cosmology completely and is ruled out by observation. At the earliest stages of galaxy formation, the energy density was of order 10^{-8} eV^4. This argument leads to the bound, if the variation is first order,

$$\left|\frac{\delta\alpha}{\alpha}\right| < 10^{-37}. \tag{5}$$

The bound is significantly weaker if the change is second order, of order 10^{-18}, but still this estimate is many orders of magnitude smaller than the variation inferred from the quasar observations. It is necessary to suppress not only the quadratic terms in the ϕ potential, but terms up to very high (roughly eighth or ninth) order, to accommodate a variation of order 10^{-4}.

As we noted, these estimates followed from an extremely conservative choice of Λ. There is every reason to believe that nature is well described by local field theory at scales below some higher Λ, possibly as high as the Planck scale M_P (this would lead to $|\delta\alpha/\alpha| < 10^{-104}$). There is no evidence that this general behavior depends on the form of the theory above the cutoff scale. In particular, corrections of this form invariably arise in string theory, in instances where one can do the calculation.

The situation is not quite so extreme in supersymmetric theories, which do generically cancel the vacuum energy. Of course in the real world supersymmetry is broken, at some scale $M_{\rm SUSY}$. The generic estimate in this situation is

$$\delta V = c\,\delta\alpha\Lambda^2 M_{\rm SUSY}^2 \,. \tag{6}$$

Supersymmetry generically cancels leading order divergences but does not do better than that. In the most optimistic scenario, a particular supersymmetric model might indeed cancel all divergences, leading to $\delta V \propto M_{\rm SUSY}^4$. Since experiment constrains $M_{\rm SUSY} > 100$ GeV, even if

$$\delta V = (\delta\alpha)^2 M_{\rm SUSY}^4 \tag{7}$$

we obtain $\delta\alpha/\alpha < 10^{-24}$.

Note that this argument would apply both to a continuous variation of α (we study this in more detail below), but also to more drastic modifications of the physics between early and late times, such as a first order phase transition. Indeed, on general grounds one expects the vacuum energy to decrease with time; in this situation a phase transition can only increase the variation of the energy.

Let us now address a possible objection to this argument. It is that normally we regard terms such as (2) as part of the cosmological constant and imagine that they are subtracted off by whatever mysterious agency resolves the cosmological constant problem.

However, if α varies, general physical principles force us to treat it effectively as a scalar field. Then, we should not do this subtraction for all values of α but only for a particular value: its asymptotic value, or perhaps its value at the current era. There are many reasons to believe this:

(i) Subtracting the cosmological constant in low energy effective field theory means fine-tuning the coefficient of a single relevant operator. Subtracting the vacuum energy for all values of a scalar field means fine-tuning the coefficients of an infinite number of relevant and irrelevant operators.

(ii) All theories of inflation and quintessence assume that the values of the potential away from the minimum of a scalar field are nonzero. Indeed, if we imagined the solution of the cosmological constant problem actually subtracted away the potential of all scalar fields, we would not even be able to implement the Higgs mechanism.

(ii) Specific proposals, e.g., [4] for explaining the mystery of the cosmological constant involve fine-tuning only the minimum value of the potential.

Thus, this objection does not invalidate our conclusion that observable variations of the fine structure constant are highly unnatural.

To complete the discussion, we now study the cosmology of a field ϕ as postulated in (1) to see what type of violations of naturalness are required to evade the bound. Besides the coupling (1), we take ϕ to be governed by the Lagrangian

$$\mathcal{L} = (\partial\phi)^2 - V(\phi). \tag{8}$$

The potential energy $V(\phi)$ will be modeled by the generic form $\mu^4 f(\phi/M)$. If we neglect Hubble friction, the natural time scale for motion of this field is M/μ^2. For plausible values of the microphysical parameters μ and M, this motion will be too rapid to fit the slow time variation of α. We therefore assume that, until very recently, the motion has been friction dominated. This requires $MH/\mu^2 \gg 1$ and $M \gg M_P/48\sqrt{\pi}$. Then

$$\delta\alpha^{-1} = \frac{4\pi\epsilon\mu^4}{3M^2}\int \frac{dt f'(\frac{\phi}{M})}{H}. \tag{9}$$

The integral is over a period in the matter dominated era of the universe. During this period, ϕ/M changes only very little, while f is a smooth function of order 1. Thus

$$\delta\alpha^{-1} \sim \frac{\epsilon\mu^4}{M^2}\frac{M_P^2}{\rho_{\rm now}}. \tag{10}$$

The energy density μ^4 can at most be of order $\rho_{\rm now}$, since otherwise we would have seen inflation rather than matter domination and the calculation would not be self-consistent. Thus we have

$$\delta\alpha^{-1} \le \frac{\epsilon M_P^2}{M^2}. \tag{11}$$

If it is of order $\rho_{\rm now}$ then ϕ is a form of quintessence [5,6]. (Fine-tuning issues in quintessence models have been considered in [7].)

At this point, one might conclude that $\mathcal{O}(1)$ values for ϵ and M/M_P could lead to the desired result. However, we still need to consider the effect of ϵ on the vacuum energy. This can be estimated along the lines discussed earlier. We assume a supersymmetric theory to obtain

$$\delta V = (\Lambda M_{\rm SUSY})^2 f_2(\epsilon\phi/M) \tag{12}$$

in terms of a smooth $\mathcal{O}(1)$ function f_2. We furthermore assume $\Lambda \ge M_{\rm SUSY}$ and then set $\Lambda = M_{\rm SUSY}$ to obtain a conservative bound.

In order that this interaction does not change our estimate of the size of the potential, we must insist that

$$\epsilon \le \frac{\sqrt{\rho_{\rm now}}}{M_{\rm SUSY}^2}, \tag{13}$$

and this gives a stringent bound on the time variation of the fine structure constant:

$$\frac{\delta\alpha}{\alpha} \le \alpha\,\frac{\rho_{\rm now}^{1/2}M_P^2}{M^2 M_{\rm SUSY}^2} \sim 10^{-28}\left(\frac{M_P}{M}\right)^2. \tag{14}$$

Thus taking $\delta\alpha \sim 10^{-4}\alpha$ is inconsistent with our initial assumptions. That is, a field with $M < 10^7$ GeV and $\mu^4 \sim \rho_{\rm now}$ would not undergo friction dominated motion during any part of the matter dominated era of the universe. Restoring consistency by requiring $M > M_P$ brings us back to the bound (5).

The essential point is the same as in the first argument, that for such a variation to be consistent with cosmology requires fine-tuning of the potential over a range of parameters. Indeed, if we want $V(\phi) \sim \rho_{\rm now}$ over the entire range $0 \leq \delta\alpha \leq 10^{-4}\alpha$, we must fine-tune away contributions from (roughly) the first ten coefficients of the Taylor expansion of the function g in (12).

A potential loophole in the argument as we just formulated it is that we assumed the coupling (1), but this precludes the interesting case that $V(\phi)$ vanishes at infinity and that the scalar is evolving towards arbitrarily large values. Many models of quintessence assume such a potential.

This point can be dealt with by generalizing the Lagrangian to

$$\mathcal{L} = g(\phi)(\partial\phi)^2 - V(\phi). \tag{15}$$

Here $g(\phi)$ is an arbitrary positive function which might be thought of as related to wave function renormalization, or better as a metric on the configuration space. Within this family of Lagrangians, we can perform arbitrary field redefinitions $\phi \rightarrow \phi'(\phi)$.

In fact, we can use this freedom to redefine ϕ so that the coupling (1) is exact; in other words define ϕ by the relation $\epsilon\phi/M = \delta\alpha^{-1}$, the variation $\delta\alpha$ around its late time asymptotic value. The possibility we missed of ϕ going to infinity becomes, after the field redefinition, the possibility that the integral $\int d\phi\, g(\phi)^{1/2}$, the invariant measure of distance in field space, could diverge as $\phi \rightarrow 0$. (We are not being perfectly general yet as we are assuming that the original relation between ϕ and $\delta\alpha$ was one to one. One can generalize further, but this does not lead to further illumination.)

Under our previous assumptions, the slow roll analysis goes through in the same way, with the equation of motion

$$\dot{\phi} = -g^{-1}(\phi)V'(\phi)$$

and corresponding modifications to (9) and other equations. However, the main point does not require any detailed analysis. It is that the new possibility, in this language that $g(\phi)$ diverges as $\phi \rightarrow 0$, would lead to a slowing down of $\dot{\alpha}$ but does not affect the essential point as seen in the first argument.

In general, the limit $\phi \gg \Lambda$, which might be thought to be problematic in effective field theory, in many cases is not. Such field redefinitions can also be used to clarify other limits; for example $M \gg M_P$, etc.

All this is not to say that scalars with potentials $V(\phi) \sim \rho_{\rm now}$ are impossible. This still requires an extreme fine-tuning, of course, but if we place suitable bounds on ϵ, leading to unobservably small $\delta\alpha/\alpha$, one can argue that one has done only two fine-tunings, of μ/M_P and of ϵ. The most natural way to accomplish the second of these would be to simply postulate a symmetry enforcing $\epsilon = 0$.

As far as we know, the only possible explanation of such small numbers would be an axionlike shift symmetry for the scalar field ϕ. General arguments rule out global continuous symmetries in theories including gravity. However, there are axionlike fields that arise from higher dimensional antisymmetric tensor gauge fields. A generic context in which one might expect small axion potentials to be generated is that of brane world models, like that of Horava and Witten [8]. There axions arise as would be gauge modes of bulk gauge fields, and potentials are generated by interaction with the boundary. In some Horava-Witten models for the real world, one can obtain axions whose potentials are generated by weak interaction instantons [9]. This can give rise to very small numbers for couplings that would be of order 1 by dimensional analysis.

This reinforces the conclusion of [10] that axions are the only known model of quintessence that might be consistent with the naturalness constraints of quantum field theory. Note, however, that [10–12] all point out various problems with the idea of axions as quintessence.

Our overall conclusion is that we do not have *any* field theoretically natural explanations for a variation of the fine structure constant as large as would be required to explain the observations of [1]. If these observations are confirmed, one will have to invent some very exotic physics to explain them.

Following the philosophy we are advocating, and again assuming that we are near a minimum of the effective potential, general arguments about minima of functions suggest that any correlated variation of couplings will yield a similar bound. In particular, the limits on the fractional variation of the electron mass and the QCD coupling will be even more severe than those we have discussed here.

The research of M. R. D. and T. B. was supported in part by DOE Grant No. DE-FG02-96ER40959, and the research of M. D. and T. B. was supported in part by DOE Grant No. DE-FG03-92ER40689.

Note added.— In this Letter, we have focused on the variation of the fine structure constant, since this is the manner in which the result of [1] is presented. However, in any model in which the variation of a field leads to variation of the fine structure constant, it is likely to lead to variation of the other constants of nature, including the electron mass and the QCD coupling, as well as (if nature is supersymmetric) the scale of supersymmetry breaking. Arguments suggesting correlated variations of several couplings have recently been made in [13].

*Email address: banks@physics.rutgers.edu
†Email address: dine@scipp.ucsc.edu
‡Email address: mrd@physics.rutgers.edu

[1] J. K. Webb, M. T. Murphy, V. V. Flambaum, V. A. Dzuba, J. D. Barrow, C. W. Churchill, J. X. Prochaska, and A. M. Wolfe, Phys. Rev. Lett. **87**, 091301 (2001).
[2] N. Kaloper and L. Susskind (unpublished).
[3] J. F. Donoghue, hep-ph/0101130.
[4] T. Banks, hep-th/0007146.
[5] P. J. E. Peebles and B. Rhatra, Phys. Rev. D **37**, 3406 (1988); R. R. Caldwell, R. Dave, and P. J. Steinhardt, Phys. Rev. Lett. **80**, 1582 (1998).
[6] C. T. Hill, D. N. Schramm, and J. N. Fry, Comments Nucl. Part. Phys. **19**, 25 (1989). A model in which an oscillating field leads to variation of cosmological parameters is discussed in C. T. Hill, P. J. Steinhardt, and M. S. Turner, Phys. Lett. B **252**, 343 (1990).
[7] C. Kolda and D. H. Lyth, Phys. Lett. B **458**, 197 (1999).
[8] P. Horava and E. Witten, Nucl. Phys. **B475**, 94 (1996); **B460**, 506 (1996).
[9] T. Banks and M. Dine, Nucl. Phys. **B479**, 173 (1996).
[10] S. M. Carroll, Phys. Rev. Lett. **81**, 3067 (1998).
[11] K. Choi, Phys. Rev. D **62**, 043509 (2000).
[12] T. Banks and M. Dine, J. High Energy Phys. **0110**, 012 (2001).
[13] P. Langacker, G. Segre, and M. J. Strassler, hep-ph/0112233.