Conformal invariance, microscopic physics, and the nature of gravitation

Jacob D. Bekenstein
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106 and Physics Department, Ben Gurion University of the Negev, Beer Sheva, Israel

Amnon Meisels
Physics Department, Ben Gurion University of the Negev, Beer Sheva, Israel
(Received 7 September 1979; revised manuscript received 9 June 1980)

The question of whether the gravitational "constant" can vary in spacetime has been among the most vexing in physics. The thrust of this paper is that the issue may be fully resolved if one accepts the principle (first proposed by Weyl and lucidly discussed by Hoyle and Narlikar) that all the fundamental equations of physics should be invariant under local (spacetime-dependent) transformations of units (principle of conformal invariance). Theoretical arguments in favor of the principle are discussed. We then show that the presently accepted dynamics for the fundamental particles and their electromagnetic, weak, and strong interactions indeed satisfy the principle. Their conformal invariance is due not least to the indispensable transformation properties of rest masses. Thus in arbitrary units each type of rest mass is a spacetime field. The principle of conformal invariance then demands conformal invariance of the dynamics of each such "mass field." If all rest-mass ratios are strictly constant there is only one mass field. Its dynamics automatically induces dynamics for gravitation. In units defined by particle masses the gravitational action is manifestly that of general relativity, a fact discovered in different guises and independently by several workers. This would seem to forbid the construction of a conformally invariant theory of gravitation with "varying gravitational constant" G. Such theories have been proposed by Dirac, and later by Canuto and coworkers, who have argued that, *a priori*, gravitational (Einstein) units are distinct from those defined by matter (atomic units). We find that to implement such distinction while simultaneously avoiding undetermined elements in the theory, one must introduce conformally invariant dynamics for gravitation and for the mass field separately. We construct this theory; it is a "varying-G theory." We then show that it is definitely ruled out by the solar-system gravitational experiments. We conclude that the principle of conformal invariance requires that gravitation be described by general relativity, and that the dimensionless gravitational constant γ be strictly constant. We also consider the possibility that gravitation, or the mass field, explicitly break conformal invariance. The corresponding theory, the theory of variable rest masses (VMT), was developed earlier from a different viewpoint. Although it predicts variability of γ, we point out that for a vast majority of cosmological models, the temporal variability of γ is well below experimental sensitivities.

I. INTRODUCTION

Ever since Dirac[1] raised the possibility that the gravitational "constant" may vary, the vexing problem of either ruling out, or else confirming temporal or spatial variations of the fundamental constants, has confronted physics. Such variations would unquestionably complicate greatly our views of nature: they would negate the equivalence principle, the touchstone of gravitation theory. But ignoring variations if they exist would unquestionably lead to our inferring a biased picture of the Universe. Hence the problem must be met face on.

The task is not as formidable as it might have been because only variations of dimensionless constants need be considered.[2,3] Thus the quantities that must be scrutinized are the dimensionless coupling constants of the electromagnetic, weak, gravitational, and strong interactions α, β, γ, and δ, respectively, and all available rest-mass ratios m_1/m_2. A variety of ingenious arguments drawing on laboratory, astronomical, geological, and geochemical data have been used to set stringent bounds on possible temporal variations of α,[4-7] β,[4,7] δ,[6-8] and m_e/m_p (Refs. 6 and 9) (e for electron, p for proton) over the past few billion years. The Eötvos-Dicke-Braginski experiments can be interpreted as strongly constraining possible spatial variations of α, β, δ, m_e/m_p, and m_n/m_p (n for neutron).[10] Nevertheless, in the absence of concrete theories of variability, these limits are not sufficient to rule out present weak variability of the constants, or even violent variations in the early Universe. This is the more true for γ; theories of γ variability abound, but the experimental bounds on it are far from stringent.[10]

One consequence of variability of the fundamental constants would be the distinction it would draw between different "natural" systems of units, of which three are of special interest: atomic (or electromagnetic) units, particle units, and Planck-Wheeler (or Einstein or gravitational) units. The atomic unit of the length is the Bohr radius $\hbar^2/e^2m_e$ and the corresponding unit of mass is twice the Rydberg mass equivalent $e^4m_e/\hbar^2c^2$. In particle units $\hbar/m_pc$ and m_p are the units of length and mass. The Planck-Wheeler unit of length is $(\hbar G/c^3)^{1/2}$, and that of mass $(c\hbar/G)^{1/2}$, where G is that coupling constant appearing in the gravitational action. Units of time are the corresponding units of length divided by c. One could also imagine hybrids of these systems. Now the conversion factors between particle and Planck-Wheeler units depend on γ. Hence, any variation of γ would make Planck-Wheeler and particle units intrinsically different (in contrast with cgs and mks units which are equivalent). Similarly, variation of α or of m_e/m_p would make atomic and particle units different.

Our purpose is to demonstrate that the issue of variability of the constant γ may be fully resolved if one accepts as a fundamental principle that all the basic equations of physics should be fully invariant under local transformations of the unit of length. Of course, it is well known that the laws of physics look the same in the cgs and in the mks systems—this is a statement of the invariance of physics under *global* units transformations. The intent here is that the invariance should still hold for all *local* units transformations—transformations with a spacetime-dependent conversion factor. Such transformations were first considered by Weyl (who called them gauge transformations)[11] and were later lucidly discussed by Dicke,[2] Hoyle and Narlikar,[12] and Hoyle.[13] Following them we call these conformal transformations (CT's) and the invariance under them conformal invariance (CI). A CT is viewed as "stretching" all lengths (because units are changed) by factors Ω which depend only on the spacetime locations of the objects in question, and as stretching all durations by exactly the same Ω's. Thus c is unaffected by CT's. If the spacetime metric $g_{\mu\nu}$ is regarded as carrying dimensions of length squared, while the coordinates are dimensionless, the effect of a CT on length and time intervals is represented by the transformation

$$g_{\mu\nu} \to g_{\mu\nu}\Omega^2 , \qquad (1)$$

where Ω is an arbitrary, positive, smooth, and dimensionless spacetime function. The coordinates are unchanged by the CT. Since Compton lengths of particles λ are like any other lengths in physics, Dicke and Hoyle and Narlikar require them to transform as

$$\lambda \to \lambda\Omega . \qquad (2)$$

This is equivalent to requiring that rest masses transform as

$$m \to m\Omega^{-1} , \qquad (3)$$

with $\hbar \to \hbar$. In this scheme a physical field with dimensions $L^\alpha T^\beta M^\gamma$ transforms by $\Omega^{\alpha+\beta-\gamma}$.

It is important to contrast CT's with the currently popular scale transformations.[14] These last are transformations like (1) together with a rescaling of fields by powers of Ω. But Compton lengths (masses) are assumed to remain untransformed. Thus scale transformations are *not* units transformations; they are active enlargements of a system whose usefulness depends on the absence of a scale of length. By contrast, CT's *are* units transformations whose very meaning derives from the existence of some scale of length in the problem.

Weyl[11] was the first to contemplate the principle of CI. Later Hoyle and Narlikar[12] presented arguments that recommend it as a guiding principle for formulating all physical theory. It is true that it does not rest on such a firm experimental basis as, say, the principle of Lorentz invariance. Nevertheless, the principle of CI is highly attractive, not least because of its similarity to the gauge principle which has so enriched contemporary physics. Global units transformations are analogous to global gauge transformations or global internal-symmetry transformations (units transformations are concerned with redefining the magnitudes of fields, the others with redefining their phases or identities). The extension of units transformations to the local level (CT's), and the requirement of CI of physical law then parallel the promotion of gauge and internal invariances to the local level by the introduction of gauge fields.[15] The fruitfulness of this last procedure for the understanding of the elementary interactions has by now been amply demonstrated.

As further support for the principle of CI one can mention the CI of Maxwell's equations and the massless Dirac equation.[12,16] Less well known is the existence of a CI version of the massless scalar field equations.[17] That rest-mass terms do *not* break CI is a point made from time to time,[12,16,18] but largely ignored. In Sec. II we review all these issues, and also show that the electromagnetic, weak, and strong interactions between elementary particle fields, as presently understood, are CI. Thus microscopic physics is CI in its entirety. A consequence of this is that in arbitrary units each type of rest mass can be a spacetime field. The principle of CI would then require that each such mass field have CI dynamics.

This point is taken up in Sec. III where we provisionally assume that all mass ratios are strictly constant. The principle of CI and standard physical restrictions then fix the dynamics of the one remaining mass field to a form which has been used in related contexts by Hoyle and Narlikar,[12] Deser,[19] Bramson,[18] Dirac,[20] and Canuto *et al.*[21] The action for the mass field can then be reinterpreted as that for gravity: It is just the general-relativistic one. There is also a strong suggestion that the cosmological constant must vanish. Hence the simplest implementation of the principle of CI to all physics requires γ to be strictly constant, and gravitation to be described by general relativity (GR).

This would seem to rule out the construction of a CI gravitational theory with varying gravitational constant G. Such theories have been proposed by Dirac[20] and by Canuto *et al.*[21] who have argued that, *a priori*, Planck-Wheeler units are different from atomic units. In Sec. IV we point out that whereas Dirac's theory does not actually succeed in embodying the required difference between units, that of Canuto *et al.* does so only at the expense of introducing an undetermined element and making the theory incomplete. Only one way suggests itself for correcting such incompleteness: to introduce CI dynamics for the mass field and for gravitation separately. We construct the most general such theory and show that it contradicts solar-system experiments and can definitely be ruled out. Hence the principle of CI and experience leave no room for γ variability; they require Planck-Wheeler and particle units to be identical.

Finally, in Sec. V we consider the possibility that CI may be explicitly broken by the mass field dynamics, or those of gravitation. The most general theory which does this, the theory of variable rest masses (VMT), has been studied earlier from another viewpoint.[22] It does allow γ variability but we point out that realistic cosmological models within the theory generally predict temporal variations of γ due to the expansion of the Universe which are well below foreseeable experimental sensitivities. Section VI summarizes our conclusions.

We employ the following conventions: metric signature +2, Greek indices are spacetime indices, Latin ones are internal indices:

$$V_{\alpha;\beta;\gamma} - V_{\alpha;\gamma;\beta} = -R_{\alpha\delta\beta\gamma}V^{\delta}, \quad R_{\beta\sigma} = R^{\alpha}_{\beta\alpha\delta}, \quad R = R_{\alpha}{}^{\alpha}.$$

II. CONFORMAL INVARIANCE OF MICROSCOPIC PHYSICS

One often hears the statement that rest masses of fields break the CI of their dynamics. If by CI one means invariance under local units transformations (CT's), nothing could be further from the truth. The mass terms in the Lagrangian density of any field is the product of λ^{-1} or λ^{-2} with the square of the field. This term must have dimensions $ML^{-1}T^{-2}$ in order that its integral over four-dimensional volume have the dimensions of relativistic action—those of $\hbar c$. Thus under a CT the mass term is multiplied by Ω^{-4}. The four-dimensional volume element is multiplied by Ω^4. Thus the contribution of the mass term to the total action is CI—the mass term does not break CI.

The key to this result is the transformation law (2) for λ. To those adept in *scale* transformations this law may seen strange. Yet, Compton lengths of particles are fundamental "metersticks" in physics. When discussing a transformation of all lengths, one has no logical choice but to demand (2). Misunderstanding on this point has led to the myth of the breaking of CI by masses.

Another reason for this myth is the belief that a traceless stress-energy tensor $T_{\alpha\beta}$ is the hallmark of a field with CI dynamics. This is supported by the examples of the electromagnetic and neutrino fields, both CI and having $T \equiv T^{\alpha}_{\alpha} = 0$, and by that of the massless Klein-Gordon field which is not CI, and has $T \neq 0$. Since mass terms contribute to T, it is usually inferred that a massive field cannot be CI.

The fallacy in this reasoning can be seen by considering a massive field F with CI action S (examples below). A general functional variation of S is

$$\delta S = \int [(\delta S/\delta F)\delta F + (\delta S/\delta m)\delta m + (\delta S/\delta g^{\mu\nu})\delta g^{\mu\nu}](-g)^{1/2} d^4x, \qquad (4)$$

where we allow for the fact that under CT's m becomes a spacetime function, so S can also be varied by functionally varying m. The field equations for F give $\delta S/\delta F = 0$. Now suppose the variation envisaged is one generated by a CT like (1)–(3) with $\Omega = 1 + \delta\Omega$, where $\delta\Omega$ is an arbitrary infinitesimal function. Thus $\delta g^{\mu\nu} = -2g^{\mu\nu}\delta\Omega$, while $\delta m = -m\delta\Omega$. Since $\delta S/\delta g^{\mu\nu} = -\frac{1}{2}T_{\mu\nu}$, it follows from the CI of S ($\delta S = 0$), and the arbitrariness of $\delta\Omega$, that

$$T = m(\delta S/\delta m). \qquad (5)$$

Thus even though the field F has CI dynamics, its $T_{\mu\nu}$ is not traceless. Therefore, the common wisdom must be rephrased to state that a field with CI dynamics is one whose T contains only mass terms, if any.

Having established that massive fields can have CI dynamics, we now review the CI invariance of the most commonly encountered fields in physics. The field equations must be formulated in curved spacetime because, even if we start in Minkowski spacetime, a general CT will make the spacetime nonflat. From this angle it is clear that the issue of CI is inseparable from that of gravitation.

A. Vector field

The action is

$$S_V = -\frac{1}{4}\int (F^{\alpha\beta}F_{\alpha\beta} + 2\lambda_V{}^{-2}A_\alpha A^\alpha)(-g)^{1/2} d^4x, \qquad (6)$$

where

$$F_{\alpha\beta}=A_{\beta,\alpha}-A_{\alpha,\beta}\,. \tag{7}$$

From $[S]=[\hbar c]$ and $[g^{\alpha\beta}]=L^{-2}$ it follows that $[A_\alpha]=M^{1/2}L^{3/2}T^{-1}$. Thus under CT's

$$A_\alpha\to A_\alpha\,,\quad F_{\alpha\beta}\to F_{\alpha\beta}\,. \tag{8}$$

Since $g^{\alpha\beta}\to\Omega^{-2}g^{\alpha\beta}$ while $(-g)^{1/2}\to\Omega^4(-g)^{1/2}$, S is seen to be CI almost trivially [recall (2)]. With $\lambda_V{}^{-1}=0$, (6) gives the dynamics of the (Maxwell) electromagnetic field. In that case (6) is invariant under a combined CT and a gauge transformation with

$$A_\alpha\to A_\alpha+\chi_{,\alpha}\,, \tag{9}$$

where χ is an arbitrary function. If $\lambda_V{}^{-1}\neq 0$ (Proca field) (6) describes the free dynamics of a vector-meson field like that of the ρ.

With a suitable modification of (7) one can describe the CI dynamics of a non-Abelian gauge field.[15] In this case one deals with a collection of n fields A^i_α $(i=1,2,\ldots,n)$. Then

$$F^i_{\alpha\beta}=A^i_{\beta,\alpha}-A^i_{\alpha,\beta}+fc_{ijk}A^j_\alpha A^k_\beta \tag{10}$$

defines the ith antisymmetric tensor field. Here c_{ijk} are the (dimensionless) structure constants of the group associated with the field, and f is a coupling constant. The action is

$$S_G=-\tfrac{1}{4}\int F^{i\alpha\beta}F^i_{\alpha\beta}(-g)^{1/2}d^4x\,. \tag{11}$$

Evidently the dimensions of A^i_α and $F^i_{\alpha\beta}$ are those of A_α and $F_{\alpha\beta}$; thus $[f]=M^{-1/2}L^{-3/2}T$ and under CT's $f\to f$. Therefore, (8) applies; the CI of S_G is then evident. S_G can represent the dynamics of the gluon fields responsible for the strong interactions which bind quarks into hadrons according to quantum chromodynamics,[23] or the dynamics of Yang-Mills fields associated with the isotopic symmetry of hadrons.[24]

To deal with the electromagnetic interaction of a Proca field B_μ, one must regard it as complex and form its field tensor $H_{\alpha\beta}$ by the minimal-coupling prescription

$$H_{\alpha\beta}=B_{\beta,\alpha}-B_{\alpha,\beta}-i\epsilon(A_\alpha B_\beta-A_\beta B_\alpha)\,, \tag{12}$$

where ϵ is the field's charge divided by $\hbar c$, and A_α the electromagnetic potential. One finds $[\epsilon]=M^{-1/2}L^{-3/2}T$ so $\epsilon\to\epsilon$ under CT's. The action is similar to (6) but formed by multiplying B_α and $H_{\alpha\beta}$ by their complex conjugates. Its CI is evident. It is also invariant under the gauge transformation

$$B_\alpha\to B_\alpha\exp(i\epsilon\chi) \tag{13}$$

together with (9).

B. Dirac field[25]

This is represented by a four-spinor field Ψ. From the ordinary Dirac matrices γ_a $(a=0,1,2,3)$ one constructs 4×4 matrix fields γ_μ transforming collectively like a vector under coordinate transformations:

$$\gamma_\mu=\lambda^a_\mu\gamma_a\,. \tag{14}$$

Here λ^a_μ are four orthogonal covariant vector fields chosen so that

$$\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu=2g_{\mu\nu}I\,, \tag{15}$$

where I is the unit matrix.

The spinor covariant derivative acting on Ψ is $\Psi_{,\nu}-\Gamma_\nu\Psi$ where the matrices Γ_ν may be chosen as

$$\Gamma_\nu=\tfrac{1}{4}\gamma_a\gamma^\mu\lambda^a_{\mu;\nu}\,. \tag{16}$$

The Dirac action is written as

$$S_D=\int\Psi^\dagger\gamma[\gamma^\nu(\Psi_{,\nu}-\Gamma_\nu\Psi)+\lambda_D{}^{-1}\Psi](-g)^{1/2}d^4x\,, \tag{17}$$

where γ is a dimensionless 4×4 matrix field satisfying

$$\gamma\gamma_\mu+\gamma_\mu^\dagger\gamma=0\,, \tag{18}$$

$$\gamma_{,\nu}+\gamma\Gamma_\nu+\Gamma_\nu^\dagger\gamma=0\,. \tag{19}$$

Comparing (14) and (15) with (1) we see that

$$\lambda^a_\mu\to\lambda^a_\mu\Omega\,,\quad\gamma_\mu\to\gamma_\mu\Omega \tag{20}$$

under CT's. It then follows from (16) that

$$\Gamma_\nu\to\Gamma_\nu+\tfrac{1}{4}(\gamma^\beta\gamma_\nu-\gamma_\nu\gamma^\beta)\Omega^{-1}\Omega_{,\beta}I\,, \tag{21}$$

where we have used $\gamma_\nu\gamma^\nu=4I$ which follows from (15). Although Γ_ν is dimensionless it does not go into itself under CT's; this need not occasion alarm since Γ_ν (like $\Gamma^\alpha_{\beta\gamma}$) is not a tensorial quantity, and is thus not measurable. It is evident from (18), (19), and (21) that we may assume

$$\gamma\to\gamma\,. \tag{22}$$

Finally, from dimensional analysis of (17) it is clear that $[\Psi]=M^{1/2}T^{-1}$, so under CT's

$$\Psi\to\Psi\Omega^{-3/2}\,. \tag{23}$$

The CI of S_D now follows from (20)–(23) if (15) is employed to commute the γ_ν in the expression for the conformal change in $\gamma^\nu\Gamma_\nu$.

The Dirac field is coupled to the electromagnetic field by the prescription

$$\Psi_{,\nu}\to\Psi_{,\nu}-i\epsilon A_\nu\Psi \tag{24}$$

applied to (17) (again ϵ is the charge divided by $\hbar c$). Then S_D is evidently invariant under combined CT's and gauge transformations analogous to (9) and (13). The S_D then represents the dynamics of the charged leptons.

The dynamics of neutrinos is given by (17) with

$\lambda_D^{-1}=0$ together with the (left-handedness) condition[25]

$$(1-\gamma^5)\Psi=0\,, \tag{25}$$

where

$$\gamma^5\equiv(i/4!)\epsilon^{\alpha\beta\gamma\delta}\gamma_\alpha\gamma_\beta\gamma_\gamma\gamma_\delta\,, \tag{26}$$

and the $\epsilon^{\alpha\beta\gamma\delta}$ is the Levi-Civita tensor. Since $\epsilon^{\alpha\beta\gamma\delta}$ includes a factor $(-g)^{-1/2}$, $\gamma^5\to\gamma^5$ under CT's, so the condition (26) is CI, and so is the whole dynamics of the neutrino.

The formalism with suitable modification can describe the dynamics of quarks in interaction with gluons (quantum chromodynamics)—the scheme for understanding hadron structure.[23] In this case Ψ is a column vector of three quark four-spinors representing the three colors. Let T^i be the eight generators of the SU_3 group which describes the symmetry under color relabeling. The T^i can be represented as dimensionless 3×3 matrices acting in "color space." Their commutation relations contain the structure constants c_{ijk} mentioned earlier. The action for the quarks is then (17) with the replacement

$$\Psi_{,\nu}\to\Psi_{,\nu}-ifT^iA^i_\nu\Psi\,, \tag{27}$$

where A^i_ν are the gluon fields mentioned earlier, and f the coupling constant defined above. The CI of the quark action is then evident. If various quark flavors are to be considered, there is a sum of actions, one for each flavor, with λ_D^{-1} different for each.[23]

An analogous treatment can be given for baryons of an isospin doublet (i.e., proton and neutron) interacting via Yang-Mills fields associated with the SU_2 isospin symmetry.[24,26] Another example of spinor fields coupled to gauge fields is found in the Weinberg-Salam unified theory of the weak and electromagnetic interactions[15,27] in which the lepton fields are coupled to four gauge fields (transforming according to $SU_2\otimes U_1$) in the manner of (27). In both examples the CI of the spinor action is established as above.

C. Scalar field

The pseudoscalar mesons are nowadays regarded as bound states of quark pairs. Nevertheless, it is still useful to consider the dynamics of a scalar field φ as representing that of a meson in some effective way. Now, the Klein-Gordon action, when generalized naively to curved spacetime, is not CI. However, as Penrose showed long ago,[17] inclusion of the scalar curvature R in the action

$$S_s=-\tfrac{1}{2}\int(\varphi_{,\alpha}\varphi^{,\alpha}+\tfrac{1}{6}R\varphi^2+\lambda_s^{-2}\varphi^2)(-g)^{1/2}d^4x \tag{28}$$

assures its CI. We see that $[\varphi]=M^{1/2}L^{1/2}T^{-1}$, so that under CT's

$$\varphi\to\varphi\Omega^{-1}\,. \tag{29}$$

One also knows that[28]

$$R\to\Omega^{-2}R-6\Omega^{-3}(-g)^{-1/2}[(-g)^{1/2}g^{\alpha\beta}\Omega_{,\alpha}]_{,\beta}\,. \tag{30}$$

A short calculation involving an integration by parts shows that the last term in (30) serves to cancel the terms involving $\Omega_{,\alpha}$ which arise from the $\varphi_{,\alpha}\varphi^{,\alpha}$ in (28). Thus the curvature scalar acts very much like a gauge field to bring about CI.

The field is coupled to electromagnetism by making it complex and applying a prescription parallel to (24). This does not affect the CI. The scalar field can also be coupled to spinor fields; an example is the Yukawa coupling

$$\int\Psi_p^\dagger\gamma\varphi_\pi\Psi_n(-g)^{1/2}d^4x \tag{31}$$

between proton, neutron, and pion fields. From (22), (23), and (29) it follows that this coupling action is CI. The same cannot be said about the once popular derivative coupling.[29] It is interesting that the principle of CI rules out this coupling which is known to be unphysical on grounds of lack of renormalizability. Another example of this selectivity of CI for good physics is the fact that the stress-energy tensor corresponding to the CI action (28) has softer divergences than those for the non-CI Klein-Gordon field.[30]

Of fundamental importance in the Weinberg-Salam theory are the Higgs scalar fields responsible for generating rest masses for those three gauge fields which represent the W (intermediate) vector bosons.[15] The Higgs fields may be represented by a two-component complex column vector Φ. Define

$$D_\mu\Phi=\Phi_{,\mu}-\tfrac{1}{2}if'B_\mu\Phi-\tfrac{1}{2}if\tau^iA^i_\mu\Phi\,, \tag{32}$$

where B_μ and A^i_μ $(i=1,2,3)$ are the gauge fields, τ^i are the 2×2 matrix generators of SU_2, and f and f' are coupling constants with dimensions $M^{-1/2}L^{-3/2}T$. The CI action can be expressed as

$$S_H=-\tfrac{1}{2}\int[(D_\mu\Phi)^\dagger D_\mu\Phi+\tfrac{1}{6}R\Phi^\dagger\Phi-\lambda_H^{-2}\Phi^\dagger\Phi+k(\Phi^\dagger\Phi)^2](-g)^{1/2}d^4x\,, \tag{33}$$

where the mass term appears with a sign opposite the conventional one,[15] and a self-interaction term with coupling constant k is included. Clearly, $[\Phi]=M^{1/2}L^{1/2}T^{-1}$ while $[k]=M^{-1}L^{-3}T^2$. Thus under CT's Φ transforms as φ in (29) while $k\to k$. It takes but one moment to verify the CI of S_H.

D. Other fields

There are well-known problems with the naive curved-spacetime formulations of spin-$\frac{3}{2}$ and spin-2 fields: inconsistency with supplementary conditions and noncausal propagation. For the massless spin-$\frac{3}{2}$ field a consistent formulation has been achieved only within the framework of supergravity theory.[31] This scheme, however, views the spin-$\frac{3}{2}$ fermion field as a component of gravitation. Our view of the nature of gravitation (Secs. III–VI) is so remote from that of supergravity that we shall have no occasion to consider the spin-$\frac{3}{2}$ field. Among the *known* particles, only the Ω hyperon and some ephemeral resonances, i.e., Δ^0, have spin $\frac{3}{2}$. These are surely best viewed as bound states of three quarks. There thus seems to be no need to consider an elementary spin-$\frac{3}{2}$ field. Therefore, the lack of a consistent CI curved-spacetime dynamics for it is not a fundamental difficulty.

Similar remarks apply to the spin-2 field. The only consistent dynamics for a spin-2 field in curved spacetime proposed thus far[32] is invariant only under very restricted CT's [i.e., Eq. (16) of Ref. 32]. However, the current view is that a spin-2 particle like the A_2 meson is a bound state of four quarks, and does not require an elementary field for its description. No problem for the principle of CI is thus apparent.

The upshot of our discussion is that the dynamics of the fundamental particles (leptons, quarks) and of the fields by which they interact (Maxwell, gluon, W, and Higgs) according to contemporary theories of the strong, weak, and electromagnetic interactions, can be expressed in manifestly CI form. Thus microscopic physics respects the principle of CI.

III. THE MASS FIELD AND GRAVITATION

CI of microscopic physics depends on the transformation laws (2) and (3) for Compton lengths and rest masses of particles. It is evident that in a general system of units or conformal frame (CF) both λ and m for each type of field will vary in spacetime. This consequence of CI was recognized early.[16] If mass ratios are not constant (see Sec. I) this means that each type of λ^{-1} becomes a distinct spacetime field—a "mass field." Each such mass field must be a dynamical field if the theory is to be complete. One cannot hold that λ^{-1} is prescribed in one CF and is determined in others by means of (2). That "original" CF simply cannot be singled out from within the theory: all CF's are equally good before the laws. Thus the theory must provide dynamics for each λ^{-1}.

Another way to the same conclusion starts with the analogy drawn in Sec. I between CI and gauge invariance. The role of a λ^{-1} in becoming a spacetime field in order to promote global units invariance to CI is much like the role of a gauge field in promoting global gauge invariance to local gauge invariance. Thus, just as a gauge field has dynamics of its own, so should each λ^{-1} be endowed with dynamics.

In what follows we assume all rest-mass ratios m_1/m_2 are strictly constant (the more general case will be considered in a future report). As mentioned in Sec. I, there is substantial support for such an assumption from astronomical observations[6,9] and from the terrestrial Eötvös-Dicke-Braginski experiments.[10] Our assumption allows us to express all inverse Compton lengths λ^{-1} as multiples (not necessarily integral) of a universal mass field ψ with $[\psi]=L^{-1}$ which transforms according to

$$\psi \to \psi\Omega^{-1}\,. \tag{34}$$

What are the dynamics of this mass field? An early discussion of rest-mass dynamics was given by Dicke[2] in his reformulation of Brans-Dicke gravitational theory.[33] In the context of CI, rest-mass dynamics was first discussed by Hoyle and Narlikar[12] and by Bramson[18] (see also Bicknell[34]). Our discussion has much in common with theirs, except we shall undertake an axiomatic approach in the search for the dynamics of ψ.

We assume (a) the dynamics of ψ is given by an action S_ψ, (b) S_ψ is coordinate invariant, (c) S_ψ is built from ψ, its derivatives, $g_{\alpha\beta}$ and its derivatives only, (d) the equation for ψ is of order no higher than second; the stress tensor for ψ does not involve third or higher derivatives of ψ, and (e) S_ψ is CI. Postulates (a) and (b) require little comment. Postulate (c) is meant to exclude from S_ψ extraneous objects, in particular, a scale of length which in our approach would be represented by a field. In Sec. VI we shall weaken this requirement. Postulate (d) is a guard against the introduction of causal anomalies (i.e., runaway solutions) into the dynamics of ψ or $g_{\alpha\beta}$. Postulate (e) is introduced in compliance with the principle of CI; if S_ψ is not CI, the dynamics of ψ will indirectly break the CI of microphysics at some level. The analogy with gauge fields leads to the same conclusion. Just as a gauge field has gauge-invariant dynamics, so the mass field should have CI dynamics.

The most general action satisfying (b)–(e) is

$$S_\psi = \tfrac{1}{2}\hbar c \int (\psi_{,\alpha}\psi^{,\alpha} + \tfrac{1}{6}R\psi^2 + l\psi^4)(-g)^{1/2}d^4x\,. \tag{35}$$

The factor $\frac{1}{2}$ is conventional, the $\hbar c$ ensures that S_ψ has dimensions of action, and l is an arbitrary dimensionless parameter. The sign of the action has been chosen with the benefit of hindsight. The action (35) (with or without l) also appears in earlier works.[18-21,30,34] It is also closely related to the dynamics of the propagator of Hoyle and Narlikar.[12] The scheme described here appears to be the field-theoretic equivalent of Hoyle and Narlikar's action-at-a-distance theory of masses and gravitation. It is also closely related to theories described by Bramson[18] and Bicknell[34] in which particles acquire their rest masses by interacting with ψ. Its relation with the other theories is more tenuous since rest masses play no central role in them. Deser[19] describes a *scale*-invariant theory, involving a scalar field, which reduces to GR in a particular frame. Callan, Coleman, and Jackiw[30] are motivated by the requirement that a scalar field have a trace-free stress-energy tensor, in modifying GR. Dirac[20] and Canuto *et al.*[21] use the action as the basis for theories of gravitation involving a variable gravitational constant (see Sec. IV).

At this point one would inquire into the form of the gravitational action of the theory. As various workers have discovered independently, one can reinterpret S_ψ as the gravitational action. This approach gives the simplest CI scheme for all physics: the total action is $S_\psi + S_M$, where S_M is the sum of material field actions discussed in Sec. II. The content of the theory is most clearly seen by expressing it in particle units; since we may write $m_p = g_p \psi$, where g_p is a dimensionless *constant*, our choice of CF amounts to having m_p constant and $\psi = \psi_0 = \text{const}$. S_M has then its standard "form"—with constant rest masses, while S_ψ becomes the Hilbert action for general relativity (GR) with $3c^3/4\pi\hbar\psi_0^2$ playing the role of G and $-3l\psi_0^2$ playing the role of cosmological constant. This "simplification" was first noticed by Deser[19] and Bramson[18] (and by Hoyle and Narlikar in the context of action-at-a-distance theory) who concluded that the theory reduces to GR in a particular CF.

This conclusion is, however, an understatement. The physical content of the theory is the same in any units; it is GR in any CF. For this theory in particle units we get

$$\gamma \equiv G m_p^2/\hbar c = 3g_p^2/4\pi . \tag{36}$$

Thus, *the simplest implementation of the principle of CI requires that gravitation be described by GR and that γ be strictly a constant.* We note that numerically $\gamma \simeq 10^{-38}$ so $g_p \simeq 10^{-19}$, and similar values apply to other particles; the coupling between massive fields and the mass field is very weak.

Another point concerns the cosmological constant. In (35) $l^{1/2}\psi$ plays the role of inverse Compton length for the mass field. That the field which determines particle rest masses should itself have rest mass seems incongruous. What is this "rest mass"? Its inverse Compton length squared is of order of the absolute value of the cosmological constant which is, observationally, less than 10^{-55} cm^{-2}.[35] Hence the rest mass of the mass field is at least 10^{38} times smaller than that of the electron. This strongly suggests that l vanishes exactly.

Our conclusions are based on the classical theory; they may well require modification in the quantum theory. Notice that the action (35) is that of a scalar field whose energy density is nominally negative. This might cause difficulties in quantization. Actually, one prescription for quantization of ψ and $g_{\mu\nu}$ leads to a quantum theory based on (35) with $l=0$ which is entirely equivalent to that based on the Hilbert action.[36] This is in harmony with our conclusions. But no complete study of the quantum ψ field interacting with massive particles via their rest masses has been carried out. Such study may reveal an instability of the ψ vacuum which would then require us to reject the present theory and recognize that some more complicated scheme describes rest masses and gravitation.

IV. IS CONFORMAL INVARIANCE COMPATIBLE WITH VARIABLE γ?

Our conclusion that CI requires γ to be constant was based on the venerable idea that no separate action need be introduced for gravitation; that of the mass field suffices.[12,18,19] This conclusion is out of line with a recent trend to formulate theories of variability of γ as theories with local units invariance. The first of these was Dirac's[20] revival of Weyl's old CI electromagnetic-gravitational theory,[11] this time as a two-metrics theory to embody the difference between atomic and gravitational units which he has always felt to be indicated by the large cosmological numbers.[1] There followed the Canuto *et al.*[21] modification of Dirac's scheme, and the recent reinterpretation[37] of Hoyle and Narlikar's CI action-at-a-distance theory[12] as a theory of variable γ. In enquiring whether some more general scheme based on the mass field could give a variable γ, it will evidently be useful to understand how the mentioned theories try to achieve this.

Dirac's theory is based on the CI action (our conventions)

$$S_\beta = \int (\beta^2 R + 6\beta_{,\alpha}\beta^{,\alpha})\sqrt{-g}\, d^4x \tag{37}$$

for a scalar *field* β which transforms as

$$\beta \to \beta\Omega^{-1} \tag{38}$$

under an arbitrary transformation of the standard of length [see (1)]. In (37) we have not included the cosmological term or the electromagnetic action.[20] Dirac's paper is ambiguous as to the dimensions of β, but the law (38) together with the requirement that S_β have units of action leave little option: $[\beta]=L^{-1}$ [or alternatively $[\beta]=M$; one must supply factors of $\hbar$ or c to (37)]. In particular, β may not be dimensionless as Dirac seems to suggest at several points. As the interaction between β and matter, Dirac proposes one which, for one particle, is represented by the CI action (again we suppress electromagnetism)

$$S_1=-m_0\int\beta\,ds\,, \tag{39}$$

where m_0 is a constant, and ds is the line element along the particle's world line.

In Dirac's theory Einstein (or gravitational) units are those in which β is constant; in these S_B becomes the Hilbert action. Dirac regards these units as distinct from atomic units. Yet the particular theory at hand hardly incorporates that distinction. In (39) it is $m_0\beta$ which plays the role of rest mass $\times c^2$—the factor multiplying the line element in the action of a point particle.[38] In view of (38) and the obvious consequence of (1), $ds\rightarrow\Omega ds$, it follows from CI that m_0 is a constant in any CF, not just in a particular one (its dimensions are those of $c\hbar$). Therefore, in the Einstein CF rest masses are also constant: gravitational units *are* particle units . Since Dirac regards α or mass ratios as constant, he identifies particle and atomic units. Thus, despite his avowed intention to write a theory which distinguishes between atomic and Einstein units, the theory at hand does not do this—it has constant γ and is evidently GR in a CI garb, essentially the theory of Deser and Bramson (see Sec. III). The failure to incorporate a distinction between the units can be traced to the use of a *single* field β to describe the characteristic length associated with gravitation, and that associated with rest masses.

The theory of Canuto *et al.* is also based on the action (37), which is viewed as an action for gravitation, and on the law (38). But now β is not regarded as a field, but as the *dimensionless* conversion factor between the units being used and Einstein units: β is the Ω that converts from the CF under consideration to the Einstein CF. The assumption is that the β from the atomic to the Einstein CF's, β_0, is not constant, which is equivalent to saying that γ varies (Canuto *et al.* do not question the constancy of α or m_e/m_p—they implicitly identify the particle and atomic CF's). A feature of the theory is that the spacetime dependence of β_0 is extrinsic; it is not determined by the equations, but must be put in by hand, presumably from observations.

The law (38) can be squared with the dimensionless character of β only when it is realized that β is *not* a field; its numerical value depends on *two* CF's, one of which is the Einstein CF. For some purposes one can think of β as an ordinary field which transforms according to (38) when one changes units of length provided one realizes that all such transformations are with reference to the Einstein CF which is the privileged CF in the theory. The presence of a privileged CF is also seen by realizing that since β is dimensionless, the integral in (37) does not have dimensions of action $[\hbar c]$ even if one supplies factors of $\hbar$ and c. One must of necessity divide it by the square of a length. This may *not* be a variable quantity in any CF, for inserting it under the integral would spoil the latter's invariance under (1) and (38). So there is a *constant* length in the theory in any CF. Whatever ones views as to the compatibility between such a thing and conformal or scale invariance, it is clear that this constant length defines a privileged CF. Thus the theory departs from the spirit, if not the letter, of the principle of CI—that physics should look the same in any system of units, and so there should not be a privileged one.

It must also be realized that because it does not determine β_0, the theory does not satisfy the minimal demand that a gravitational theory should be complete in order to be considered viable.[10] In effect the variability of γ is put in, not demanded by the theory. For all these reasons we do not find the theory of Canuto *et al.* convincing as a theory of variable γ which also incorporates the idea of CI.

The Hoyle-Narlikar CI gravitational theory was originally interpreted as equivalent to GR within each "domain" of the Universe.[12,13] A new interpretation has been given recently by Canuto and Narlikar[37] in which it is regarded as a framework for γ variability in the context of cosmology. This interpretation is *not* required by the equations. In Hoyle-Narlikar theory rest masses are also regarded as proportional to a universal mass field or function m; they have the form λm. For an isotropic expanding universe in which t is the cosmic time and n the particle density, the equations only require that *in units for which m is constant*, $\lambda n\propto t^{-2}$. These units can be identified as gravitational; in them the gravitational equations look like Einstein's equations (constant G). Hence $\gamma\propto\lambda^2$. Canuto and Narlikar *postulate* that $\lambda\propto t^{-1/3}$, thus obtaining varying γ. The price is that in a *comoving* volume element the number of particles varies as $t^{1/3}$: matter is not conserved. Had they made the more straightforward choice $\lambda=$const there would have been conservation of matter. And, of course, then $\gamma=$const. Since the equations do not determine λ, it would seem most appropriate to avoid arbitrary choices and to take it as constant, especially when this is the only choice which respects the conservation law. Thus, no strong evidence for the belief that the theory of Hoyle and Narlikar implies γ variability exists. The variability must be put in by hand.

Evidently, then, within the framework of genuinely CI theories involving a single dynamical action for a mass field and gravitation, γ must be strictly constant. We must now investigate the possibility that separating the issue of mass field dynamics from that of gravitational dynamics may lead to a viable CI theory of variable γ. How can one build an action S_G for gravitation which does not include the mass field ψ? Out of $g_{\alpha\beta}$ and its derivatives one can build CI scalars[11] (i.e., the square of the Weyl tensor), but they all are quadratic in the curvature. They would thus lead to fourth-order gravitational equations which would then lead to causal pathologies[39] or problems of negative energies.[40] A CI action which only involves the curvature linearly can be built only with the aid of a scalar field φ as compensator for the "bad" transformation properties of the curvature under CT's. In fact, if one defines S_G by very broad postulates paralleling postulates (b)–(e) of Sec. III, the only possibility is

$$S_G = \tfrac{1}{2}\hbar c \int (\varphi_{,\alpha}\varphi^{,\alpha} + \tfrac{1}{6}R\varphi^2 + \lambda\varphi^4)(-g)^{1/2}d^4x, \quad (40)$$

where λ is a dimensionless constant. The factor $\frac{1}{2}$ is conventional, and $\hbar c$ appears because we define $[\varphi] = L^{-1}$. Evidently, $\varphi \to \varphi\Omega^{-1}$ under CT's. The φ is directly related to the characteristic length of gravitation.

The full CI theory is now described by the action $S_G - S_\psi + S_I + S_M$, where S_ψ is defined by (35), S_I is the CI interaction term between φ and ψ, and S_M is the matter action which includes ψ (in all rest masses), but not φ. The sign of S_G has been chosen positive in order that a positive gravitational constant emerge in gravitational units (φ = const) for which S_G looks like the Hilbert action. We choose the sign with which S_ψ enters into the total action as negative so that ψ will be a field with positive energy.[41]

One conceivable CI S_I would be

$$S_I' = \hbar c A \int (\varphi_{,\alpha}\psi^{,\alpha} + \tfrac{1}{6}R\varphi\psi)(-g)^{1/2}d^4x, \quad (41)$$

where A is a dimensionless real constant. However, one can combine S_I' with the analogous terms in $S_G - S_\psi$ to get, apart from the cosmological terms proportional to l and λ, the S_G with $\varphi + A\psi$ replacing φ minus the S_ψ with $(1+A^2)^{1/2}\psi$ replacing ψ. One can even get the l and λ terms by including in (41) an appropriate quartic polynomial in φ and ψ; this is implicitly included in (42) below. Since φ does not appear in matter, $\varphi + A\psi$ is just as good as it for the role of compensator field (it transforms in the right way). Also one can use $(1+A^2)^{1/2}\psi$ as a mass field if one divides all the proportionality constants between rest masses and the mass field by $(1+A^2)^{1/2}$. Hence, inclusion of S_I' does not add anything to the physics. The only other S_I which is CI and does not introduce derivatives higher than the second in any of the field equations is

$$S_I = \tfrac{1}{2}\hbar c \int \psi^4 F(\varphi/\psi)(-g)^{1/2}d^4x, \quad (42)$$

where $F(x)$ is an arbitrary real function.

To find the physical content of the theory, we pass to particle units which amounts to setting $\psi = \psi_0$ = constant. S_M then takes on its standard form with constant rest masses while $S_G - S_\psi + S_I$ may be written as

$$S = c^4(16\pi G)^{-1} \int [\phi R - \omega(\phi)\phi^{-1}\phi_{,\alpha}\phi^{,\alpha} + \Lambda(\phi)](-g)^{1/2}d^4x, \quad (43)$$

where

$$\omega(\phi) = -\tfrac{3}{2}\phi(1+\phi)^{-1}, \quad (44)$$

$$\Lambda(\phi) = 6\psi_0^2[F(\varphi/\psi_0) + \lambda(\varphi/\psi_0)^4 - l], \quad (45)$$

$$\phi = \varphi^2\psi_0^{-2} - 1. \quad (46)$$

As in Sec. III we have defined G as $3c^3/4\pi\hbar\psi_0^2$. Thus the theory is a scalar-tensor theory with a cosmological *function* Λ.[42] For such a theory the local Newtonian gravitational constant (in particle units) is given by

$$G_N = \frac{G}{\phi}\,\frac{4+2\omega}{3+2\omega} = \frac{G}{3}\,\frac{\phi+4}{\phi}, \quad (47)$$

where the second equality applies to the specific theory being considered here. Evidently, if one defines γ in terms of G_N, the theory is a CI variable-γ theory.

Is the theory viable? The first requirement for this is that $G_N > 0$, at least in the solar system. This can be true only if the value ϕ_0 of ϕ at large distance from the sun is positive [by (46) $\phi > -1$]. Then the solar-system value of ω, $\omega(\phi_0)$, is between $-\frac{3}{2}$ and 0. By contrast, the current very conservative limit on ω of a scalar-tensor theory from the solar-system experiments is[10] $|\omega| \gtrsim 30$ (and more optimistically one believes today that $|\omega| > 200$). Hence, regardless of the choice of l, λ and F, the theory just described is strongly ruled out by experiment.

One could try to build more complicated CI theories, for example, by using more scalar fields, or a vector field in conjunction with a scalar one. No physical basis for introducing these extra fields exists. Also, the previous example suggests the theories would find it difficult to meet experimental constraints. In general, complicated many-field gravitational theories have fared badly in the confrontation with experiment.[10] So without belaboring the issue, one can state with some confidence that *there exists no viable and complete CI theory of gravitation with variable* γ.

V. BROKEN CONFORMAL INVARIANCE AND VARIABLE γ

If one's belief in the principle of CI is strong, one would now reach a negative answer to the question first raised by Dirac. We, however, think it more prudent to also consider the alternative possibility that gravitation and the mass field explicitly break CI, and that as a result γ actually varies. In this manner the final decision as to the correct gravitational theory is taken out of the province of "pure thought" or esthetic considerations, and the role of experiment as final arbiter is recognized all along.

The mass field is still defined by microphysics, but we must now consider non-CI dynamics for it. In the absence of CI we must specify the units in which the action has the form we postulate. We are first tempted to use particle units. In these ψ = const, so only $g_{\alpha\beta}$ is available for constructing the action. The only physically reasonable candidate is the Hilbert action with a cosmological term. Thus we find gravitation to be described by GR with a constant γ. Thus particle units are not the appropriate ones for our task. Of the remaining infinity of CF's, one is singled out by the physics: that whose unit of length is the characteristic length defined by gravitation, namely, the Planck-Wheeler length L_{PW}. These Planck-Wheeler units are not *a priori* identical to particle units. Let us formulate S_ψ in them by adopting postulates (a)–(d) of Sec. III (but now we allow L_{PW} to enter the theory).

At first sight the most general action allowed is (35) with the $\frac{1}{6}$ replaced by a general parameter q.[43] (We take the action with negative sign to avoid negative energies and drop the cosmological term.) But, in fact, we can multiply each term in (35) by a different function of $L_{PW}\psi$ without upsetting the postulates. [We could not do this in Sec. III because L_{PW} was explicitly excluded by postulate (c).] The functions, however, are not arbitrary for the following reason. Inverse Compton lengths are certain multiples of ψ. But we can think of these multiples as being, say, seven times smaller if we regard ψ as seven times larger. Thus multiplying ψ by a constant should not affect its dynamics. We can easily see that this will be true only if both functions are the same (real) power s of $L_{PW}\psi$, for then the constant multiplying ψ can be absorbed into the unimportant coefficient of the action. Thus by defining $\overline{\psi}= L_{PW}^{s/2}\psi^{s/2+1}$ we find that

$$S_\psi = -\tfrac{1}{2}\hbar c \int (\overline{\psi}_{,\alpha}\overline{\psi}^{,\alpha} + qR\overline{\psi}^2)(-g)^{1/2}d^4x\,. \qquad (48)$$

Let us first assume that (48) also gives the gravitational dynamics. To visualize them we now pass to particle units. An appropriate conformal factor for the transformation is $\Omega = \psi L_{PW}$ which will make masses constant. Thus we replace $g_{\mu\nu}$ in (48) by $\tilde{g}_{\mu\nu}f^{-r}$, where $f \equiv (\overline{\psi}L_{PW})^2$, $\tilde{g}_{\mu\nu}$ is the metric in particle units, and $r = 2(s+2)^{-1}$. We also express R in terms of the $\tilde{R}$ built out of $\tilde{g}_{\mu\nu}$; $\tilde{R}$ is given by the right-hand side of (30). (However, we do not bother to express $\overline{\psi}$ in terms of its particle units equivalent because we only want to focus on the dynamics of $\tilde{g}_{\mu\nu}$.) After an integration by parts we get

$$S_\psi = -\tfrac{1}{2}\hbar c \int L_{PW}^{-2}[\,qf^{1-r}\tilde{R} + \tfrac{1}{4}(1 - 12qr + 6qr^2)f^{-1-r}f_{,\alpha}f_{,\beta}\tilde{g}^{\alpha\beta}] \times(-\tilde{g})^{1/2}d^4x\,. \qquad (49)$$

This is not familiar, but define

$$\phi = -8\pi q f^{1-r}\,, \qquad (50)$$

$$\omega = \tfrac{1}{4}(12qr - 6qr^2 - 1)q^{-1}(1-r)^{-2}\,. \qquad (51)$$

Then by renaming $\tilde{g}_{\alpha\beta}g_{\alpha\beta}$ (49) can be written in the form (43) with ω constant. We recognize this as the action for the Brans-Dicke theory of gravitation (BDT).[33]

Thus, somewhat suprisingly, the explicit breaking of CI of S_ψ leads inescapably to BDT. By its original construction[33] BDT is evidentally a variable-G theory (in particle units). Its generalizations, the scalar-tensor theories,[42] involve a variable coupling constant ω and are thus not in general pure variable-G theories.[44] In fact, BDT can be viewed as the most general theory of variability of the fundamental constant G. BDT has lately fallen into disrepute because of the large values of ω required to make its predictions conform with solar-system experiments ($\omega > 200$). Our way of deriving it here shows this "shortcoming" to be fictitious. The fundamental parameter is not ω, but the CI-breaking indices $q - \frac{1}{6}$ and s (for $q = \frac{1}{6}$, $s = 0$ we have CI). For illustration consider the moderate values $s = 0.2$ ($r = 0.9$) and $q = -0.05$. We get $\phi > 0$ as required by the interpetation of ϕ^{-1} as the gravitational constant, and $\omega = 785$ in agreement with all experiments. Looked at this way BDT is perfectly viable and not expected to depart much in its predictions from GR.

Thus far our way to break CI is not the most general possible. We could also introduce an explicit non-CI action for gravitation. In Sec. III we refrained from such addition primarily because it would break CI. Here we free ourselves from this inhibition. Adding to (48) the Hilbert action we recover the VMT theory proposed by one of us[22] as the most general theory of variable rest masses (in Planck-Wheeler units) which is consistent with Einstein's equations. In particle units VMT has the form of a scalar-tensor theory with a special (but variable) ω. It is thus too general to be a theory of the variability of the fundamental constant G. In fact, because ω is variable the relation between the scalar field ϕ and G is unclear.[44] However, VMT's Newtonian approximation[22] defines a variable *Newtonian* gravitational constant G_N. Thus VMT is a theory of the variability of $\gamma \equiv G_N m^2/\hbar c$; we have argued earlier[44,45] that it is the most general such theory. (The general scalar-tensor theories cannot be characterized as neatly.[44])

In previous work,[22,44] we demonstrated that VMT's predictions for solar-system experiments, neutron stars, and black holes are in close accord with those of GR for $r < 0$ and $q > 0$. Thus, for $r < 0$ and $q > 0$ VMT is a viable theory.[46] (The case $r = 1$, $q = \frac{1}{6}$ of VMT is actually the $\psi\varphi$ theory of Sec. IV.) The agreement with GR improves all the time as the Universe expands.[47] Yet VMT has a supreme advantage over GR: it has nonsingular cosmological solutions which, at late times, are indistinguishable from those of the GR which describe the present Universe so well. In VMT models of our expanding Universe, γ decreases due to the expansion, but in the vast majority of models the *present* time scale of variation is some orders of magnitude longer than the Hubble time scale.[47] Thus Dirac's specific proposal[1] for variation on the Hubble time scale does not receive support from the VMT. By contrast many VMT models predict very large overall variation of γ from the start of the expansion till today, allowing for a semiquantitative explanation *a la* Dirac of the (present) large value of γ^{-1}. Thus if one shares Dirac's confidence that the numerical agreement between the "large numbers" is not a coincidence, one cannot help seeing this model result as a suggestion that γ is variable in nature, and that gravitation explicitly breaks conformal invariance.

VI. CONCLUSIONS

One conclusion is that microphysics is CI: the dynamics of fermions, mesons and their strong, weak, and electromagnetic interactions, as currently understood, are local units invariant. Rest masses do *not* break CI as interpreted; rather, they define a mass field with important physical implications.

A second conclusion is that mass field dynamics may be reinterpreted as gravitational dynamics. If one follows the example of microphysics and postulates CI dynamics for the mass field, one has no choice but to describe gravitation with general relativity, and to accept the absolute constancy of γ. Attempts to circumvent this logic fail because they introduce arbitrary elements into the theory, or violate in spirit the principle of CI, or run into contradictions with experience. Thus variation of γ can arise only if gravitation and the mass field do not themselves respect CI.

The third conclusion is that allowing for an explicit controlled breaking of CI in mass field and gravitational dynamics leads to precisely those gravitational theories which are pure variable-γ theories: the Brans-Dicke theory and the VMT. The first is viable for modest CI-breaking indices, the second for wide ranges of these indices. The VMT predicts, in the framework of nonsingular cosmological models, that γ varies only weakly today, but decreased very strongly in early stages of the Universe. Thus one has at hand an explanation of the large-numbers puzzle which does not imply that general relativity is a bad description for the Universe today.

The last conclusion is that since constancy of γ and CI of gravitation are two faces of the same coin, one would do well to interpret any experimental constraints on γ variability in terms of what they imply about CI or lack thereof in gravitation's dynamics.

ACKNOWLEDGMENTS

We thank S. Malin, G. Horowitz, U. Ben-Yaacov, L. Smolin, W. Unruh, and K. Kuchař for stimulating conversations and constructive criticism. JDB thanks Professor B. DeWitt and Professor W. Kohn for hospitality at the Institute for Theoretical Physics. The research of JDB was supported in part by NSF Grant No. PHY77-27084.

[1] P. A. M. Dirac, Proc. Soc. London A165, 199 (1938).
[2] R. H. Dicke, Phys. Rev. 125, 2163 (1962).
[3] For an example of the fallacy in ruling our variations of *dimensional* constants, see J. D. Bekenstein, Comments Astrophys. 8, 89 (1979).
[4] For a review see F. J. Dyson, in *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge University Press, Cambridge, 1972).

[5]F. J. Dyson, Phys. Rev. Lett. 19, 1291 (1967); A. Peres, *ibid.* 19, 1293 (1967); J. N. Bahcall and M. Schmidt, *ibid.* 19, 1294 (1967); J. P. Turneaure and S. R. Stein, in *Atomic Masses and Fundamental Constants*, edited by J. H. Sanders and A. H. Wapstra (Plenum, New York, 1976).

[6]A. M. Wolfe, R. L. Brown, and M. S. Roberts, Phys. Rev. Lett. 37, 179 (1976).

[7]A. I. Shlyakhter, Nature 264, 340 (1976).

[8]P. C. W. Davies, J. Phys. A 5, 1296 (1972).

[9]B. E. J. Pagel, Mon. Not. R. Astron. Soc. 179, 81P (1977).

[10]See C. Will, in *General Reativity*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979).

[11]H. Weyl, *Space-Time-Matter* (Dover, New York, 1952).

[12]F. Hoyle and J. V. Narlikar, *Action-at-a-Distance in Physics and Cosmology* (Freeman, San Francisco, 1974), and in *Cosmology, Fusion and other Matters*, edited by F. Reines (Colorado Associated Universities Press, Boulder, 1972).

[13]F. Hoyle, Astrophys. J. 196, 661 (1975).

[14]R. Jackiw, Phys. Today 25, 23 (1972).

[15]E. S. Abers and B. W. Lee, Phys. Rep. 9C, 1 (1973).

[16]T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. 34, 442 (1962).

[17]R. Penrose, Proc. R. Soc. London A284, 159 (1965).

[18]B. D. Bramson, Phys. Lett. 47A, 431 (1974).

[19]S. Deser, Ann. Phys. (N.Y.) 59, 248 (1970).

[20]P. A. M. Dirac, Proc. Soc. London A333, 403 (1973).

[21]V. Canuto, P. J. Adams, H. S. Hsieh, and E. Tsiang, Phys. Rev. D 16, 1643 (1977).

[22]J. D. Bekenstein, Phys. Rev. D 15, 1458 (1977).

[23]For a review see A. I. Vainshtein *et al.*, Usp. Fiz. Nauk. 123, 217 (1977) [Sov. Phys-Usp. 20, 796 (1977)].

[24]C. N. Yang and R. Mills, Phys. Rev. 96, 191 (1954).

[25]For the curved-spacetime formulation see D. Brill and J. A. Wheeler, Rev. Mod. Phys. 29, 465 (1957), or the following reference.

[26]B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).

[27]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity* (*Nobel Symposium No. 8*), edited by N. Svartholm (Almqvist and Wiksell, Stockholm, 1968).

[28]J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 139. Our convention for $R_{\mu\nu}$ is opposite in sign to that of Synge.

[29]S. S. Schweber, H. A. Bethe, and F. de Hoffman, *Mesons and Fields* (Row, Peterson, Evanston, 1955).

[30]C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) 59, 42 (1970).

[31]S. Deser and B. Zumino, Phys. Lett. 62B, 335 (1976).

[32]L. P. Grishchuk and V. M. Yudin, J. Math. Phys. 21, 1168 (1980).

[33]C. Brans and R. H. Dicke, Phys. Rev. 124, 925 (1961).

[34]G. V. Bicknell, J. Phys. A 9, 1077 (1976).

[35]Ya. B. Zel'dovich and I. D. Novikov, *Relativistic Astrophysics* (Chicago University Press, Chicago, 1971), Vol. I., p. 28.

[36]E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. 73B, 209 (1978).

[37]V. M. Canuto and J. V. Narlikar, Astrophys. J. 236, 6 (1980).

[38]L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Pergamon, Oxford, 1975), p. 25.

[39]For a concrete example see G. Horowitz and R. Wald, Phys. Rev. D 17, 414 (1978).

[40]B. DeWitt, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979).

[41]A two-scalar-fields theory has also been considered in another context by Gerald Horowitz from whom we first heard about one possible coupling between the fields.

[42]See, for example, C. Will, in *Experimental Gravitation*, edited by B. Bertotti (Academic, New York, 1974).

[43]Other types of contributions to S_ψ are ruled out for various reasons: Terms involving powers of R, $R_{\alpha\beta}$, and $R_{\alpha\beta\gamma\delta}$ give gravitational equations of fourth order. Terms like $R^{\alpha\beta}\psi_{,\alpha}\psi_{,\beta}$ contribute third-order derivatives of ψ to the source of gravitation, and third-order derivatives of $g_{\alpha\beta}$ to the equation for ψ, thus precluding the formulation of the Cauchy problem.

[44]J. D. Bekenstein and A. Meisels, Phys. Rev. D 18, 4378 (1978).

[45]A. Meisels, Ben Gurion University Dissertation, 1979 (unpublished).

[46]See also Ref. 10 where, in addition, VMT's predictions for gravitational radiation are summarized.

[47]J. D. Bekenstein and A. Meisels, Astrophys. J. 237, 342 (1980).