Gauge Gravity

and

the Unification of Natural Forces

Chuang Liu

1. Introduction.[1] If we ask what reality would be if the current theories of physics are true or approximately true or, at least, on the right track, an answer from the physics community may roughly run as follows. The physical world -- in the broadest sense of 'physical' -- consists of matter in spacetime whose parts interact with one another via force-fields[2]. It seems likely that all objects in bulk are composed of a few species of elementary particles which affect each other's behavior through one or more of the four fundamental force-fields (or interactions), viz. gravity (= the gravitational field), electromagnetism, the weak interaction, and the strong interaction. Gravity is identified with spacetime, if Einstein's general theory of relativity (GR) is true or approximately true, and the other three fields are gauge fields, if the standard model of high-energy physics is true or approximately true.[3] Because of the wave-particle duality in quantum theory, elementary particles are also fields (i.e. matter-fields) and force-fields are also particles (e.g. photons and gluons), but they still belong to markedly distinct categories.

There is nothing inconsistent or intrinsically wrong about this world picture, but the following puzzle arises for the force-fields in the form of a dilemma. Either gravity is one of the fundamental forces in nature or it is not. If it is, it then makes sense to pursue the dream of a unified field theory -- a theory which provides, or at least suggests, an ontology in which the four force-fields are different *manifestations* of a single field or substance. However, according to our best theories today, gravity alone -- whether in its classical outfit or when it is eventually quantized -- is identifiable with the 4-dimensional[4] spacetime. This implies many asymmetries among the relations between gravity and the other three

force-fields. For instance, the other three interact to one another in the background of spacetime, while their interaction with gravity is nothing more than their being in spacetime! Therefore, many are justified to say that gravitational force is no force at all; it is nothing more than the spacetime which turns out to have dynamical properties. Hence, it may be better to exclude gravity from the category of fundamental forces of nature; but if so, the pursuit of including it in the grand unified field theory -- which has ranked so high on the agenda of physics for several generations -- would not make much sense.

The dilemma is resolved if one can reduce the other three forces to spacetime as well. There are then no forces in the traditional senses of the word (more on this later), and our current theories of force-fields will be abandoned just as the theory of phlogiston was. This, as we all know, is the type of grand unification that Einstein envisaged and spent most of his career pursuing (cf. Pais 1982; Vizgin 1994). The failure of Einstein's (and others') effort in finding such a unification has put a damper on, if not altogether stopped, the pursuit in this direction.[5]

Partly because of such an asymmetry, several philosophers of science recently raised questions about the rationale of regarding gravity as a gauge field. Maudlin (1996) argues,

> there seems to be a fundamental incompatibility ... between the basic approaches of gauge field theories and GTR. ... Objects couple to the [gauge] field only if they interact through some charge that serves as a coupling constant. ... But according to GTR, *gravity simply is not a force*. ... Particles do not couple to the gravitational field, they simply exist in space-time. (p. 143, my italics)

Therefore, even if there is a unified theory with $SU(3) \otimes SU(2) \otimes U(1)$ for the three gauge fields, it is questionable whether we should expect gravity as the next stop on the ride. Weinstein (1999) observes that even though there is a sense in which the diffeomorphism group -- $Diff(R^4)$ -- of GR is a kind of gauge group (see Wald 1984), the gauge groups for the gauge fields are so radically different from it that to think of GR as a gauge field

theory would be misleading. Callender and Huggett (1999) pose the question, 'Why quantize the gravitational field?' in the title of their paper. Besides pointing out many seemingly insurmountable conceptual difficulties in the quantization of gravitational field, they explode the myth that the quantization of gravity is required by the very consistency of physical theories. Even if one adopts the canonical formulation of quantum gravity from which the theory can be viewed as a gauge-field theory, one still runs into two types of problems. First, the canonical quantum gravity is not the same kind of gauge field as the other force-fields in the program. Second, there are grave problems in canonical quantum gravity involving (a) the nature of observables and (b) the very possibility of time and change, which are not present in quantum gauge-field theories of the other forces; for details, see also Belot and Earman (1999) and the references therein.

One can always take refuge in the fact that it is ultimately an empirical question as to whether and how a grand unification including gravity can be achieved. Therefore, let us wait and see. However, given that so much philosophical consideration usually goes into theoretical projects of such depth in any area of science, it would be significant if we can see clearly from a philosophical (and foundational) perspective what seems reasonable to expect and what seems not. While carrying out a full investigation of all the possibilities in the logical space for the question of the unification of force-fields needs the length of a monograph if not more, I shall focus my attention on the following possibility (which is diametrical to Einstein's original conception of unification). Could GR be modified or replaced by another theory so that gravity can be made to join the other force-fields? Weinstein is right in arguing that GR in its original formulation is no gauge field of the Yang-Mills type, it does not, however, mean that a gauge theory of gravity -- which differs from GR -- is not possible, which both meet the experimental findings for gravity and is a genuine gauge field.[6]

I shall explore this possibility by giving in section **5** a review of the various proposals for a gauge theory of gravity, which represent a valiant attempt of the gauge-field

program to integrate gravity. But before I do that, I must first examine the concept of unification of force-fields and the gauge-field program in which a unification of all force-fields of the Yang-Mills (YM) type may be achieved. There is a sense in which the four force-fields are already unified: they are effects of localized symmetries of various 'dynamical spaces'; but the difference between the 'internal' (for gauge fields) and the 'external' (for gravity) symmetries separates gravity from the rest. This crucial difference will become clear in section **5**.

      2. Unification and symmetry. Given that the conception of unification is so essential in physics, I shall first discuss its meanings in general. While a historical study is beyond the scope of this paper, I shall nevertheless identify some historical anchors for the meanings I discuss below. More subtle differences will emerge later.

      While some unification in physics involve reductions of theories at the phenomenal level to ones at the structural level, such as those in solid state physics, others involve uniting previously separate theories at the same level into a whole, such as the theory of energy conservation and transformation. Sometimes, a unification implies a reduction of the number of phenomena or types of objects/substances, while other times it is no more than a discovery of a theoretical equation that replaces several distinct equations for different phenomena. Let me begin to make these more precise. I shall not engage here the various problems with the notion of reduction; for our purpose, it suffices if we understand 'A is reducible to B' as that A can be accounted for by B -- e.g. is derivable with bridge principles from B, if A and B are theories, or that the theory of A can be accounted for by the theory of B, if A and B are phenomena or types of objects/substances.

[I]   A unification via a reduction of the phenomenal to the (underlying) structural, which is a case of unification just in case the theory of the structural has a broader

scope than the phenomenal. (Otherwise, it is simply a case of reduction without unification.)

When Maxwell realized that light rays are electromagnetic waves, a case of such a reductive unification was achieved. The macro-micro reductions in thermostatistical physics are also examples.

[II]   A unification via an integration -- not a simple conjunction -- of separate theories into a single theory, which accounts for the integration (or the interdependence) of the phenomena.

Maxwell's unified theory of electromagnetism is a case for this (see, Maudlin 1996, p. 131). As Maudlin pointed out, Maxwell's electromagnetism is no mere juxtaposition of the theories of electricity and magnetism but a single theory which explains how one gives rise to the other in the phenomena. However, the two fields remain as separate substances. In the same sense, the discovery of the conservation and transformation of two or more forms of energy -- not the postulation of the general law of conservation of energy -- is also a unification. The electromagnetic energy, for instance, remains distinct from the mechanical energy, while the theory accounts for how one may be transformed into the other.

[III]   A unification via a replacement of theories as in [II] and a corresponding ontological reduction of objects/substances.

Einstein's special theory of relativity (SR) turns Maxwell's electromagnetism into a unification of this sort (see. Maudlin 1996), because SR renders the separate appearance of the electric or the magnetic field purely a choice of inertial reference frames; and since SR completely removes the privilege of any particular inertial frame, any residual independence

of the electric or the magnetic field is removed as well. The Minkowskian unification of space and time into a 4-dimensional spacetime may be regard as another example. Note that the same does not happen in the energy conservation case: there does not appear to be any way in which we can view different forms of energy as the results of frame or perspective switching.

[IV]   A unification via an identification of universal axioms or principles from which different theories may be derivable with the help of constitutive laws and facts.

The canonical -- the Lagrangian or the Hamiltonian -- formulation of dynamics is a typical example for this (see, Dirac 1957, pp. 48-60). The unification appears to be purely formal; in other words, the contents of the Lagrangian or the Hamiltonian formulation of different theories can be very different, not even having to share the same vocabulary. On the other hand, the Hamiltonian does represent the total energy of a system, and the universal applicability of the Hamiltonian formulation may be seen as a manifestation of the universal conservation and transformation of energy. But even if this last point is true, the formulation by no means suggests any ontological reduction as in the senses of [I] and [III].

Given our conception of the general categorical structures of physical reality, which has been more or less the same since Maxwell's field theory of electromagnetism, the above can be translated into the following ideals or goals of unification in physics.

(i)   An ontological reduction of everything, matter, radiation, and forces, to a single substance in spacetime (or space/time[7]).

A grand vision involving a unification of [I] and [III], (which may be weakened by envisaging more than one fundamental substances (in spacetime) as the ultimate reducers).

Historically, this was the unificational dream of the electromagnetic program which dominated theoretical physics at the turn of the 20th century. We see names such as Larmor, FitzGerald, Lorentz, Wien, Minkowski, Mie, Abraham, and Nordström associated with the program, and it lasted even beyond the advent of Einstein's general relativity (see, Whittaker 1953; McCormmach 1970; Liu 1994a; Corry 1999).

(ii)   An ontological reduction of everything to the geometrical properties of spacetime.

As a much stronger vision than (i), it was evidently in Hilbert's mind when he tried to combine Einstein's GR (i.e. the *Entwurf* version of it) and Mie's theory of electromagnetism (see, Hilbert, 1924; Corry 1999). It was not a widely shared ideal and it is still not clear to me what kind of ontological reduction it might have entailed even if Hilbert had succeeded in his program -- an interesting investigation which I shall leave for another occasion.

(iii)   An ontological reduction of the fundamental force fields to the geometrical properties of spacetime.

This is the ideal that Einstein spent the rest of his life pursuing after he obtained his general theory of relativity (GR) (see, Pais 1982, part III; Vizgin 1994). It is also the ideal that inspired Weyl's work in gauge-field (Weyl 1918, 1919) and, indirectly or in a different guise, the guiding spirit of the gauge-field program (see, O'Raifeartaigh 1997).

(iv)   A reduction of laws and concepts of matter and force-fields to a single theoretical system, whose axioms and schemes of derivation are general and simple.

A weaker vision than (i) and (ii) involving unifications of [II] and/or [IV]. A pure form of this ideal can be found in Planck's work for a theory of 'General Dynamics' (see, Planck 1908a,b; Goldberg 1976; Liu 1994a), in which he explicitly challenged the electromagnetic program by proposing a unification of physics under two of what he regarded as the most general principles: the least action principle and the principle of SR. We shall see in a moment that is vision is also implicit in the gauge-field program.

In addition, there is a special kind of unifications in physics, some of which Maudlin (1996) call 'perfect' unifications, which I may also described as 'rare' or 'profound.' Maudlin's examples are the unified theories of electromagnetism and gravity in SR and GR respectively. Both are unifications as the result of some symmetry or equivalence: in the former, it is the symmetry or equivalence of inertial reference frames (e.g. the equivalence of observing a magnet moving (uniformly) inside a solenoid in the co-moving frame of the magnet with observing it in the co-moving frame of the solenoid); and in the latter, it is the symmetry or equivalence of local arbitrary frames (e.g. the equivalence of observing a system locally 'free-falling' in a constant gravitational field and observing it in an inertial frame). Symmetry is usually represented in physics as an invariance under a group of transformations; and therefore, it is usually taken to signify a lack of physically real distinction -- physically indistinguishable -- among the states so transformable to one another (hence their equivalence in this sense) (see, Wigner 1976, pp. 3-109; Kosso 1999). Other examples of this kind may include the original theory of Yang and Mills for the strong interaction, in which nucleons are indistinguishable under the SU(2) isospin transformations, and QCD, in which quarks are indistinguishable under the SU(3) color transformations. As I mentioned earlier, these are unifications of type [III], and when translated into an ideal for all fundamental force-fields, we have (iii), which together with the advent of quantum theory inspired the gauge-field program. We shall see later whether any unification in that program lives up to this standard.

      3. Unification before the gauge-field program. Even though it may well be true, as will be explained later, that GR, as is, is not a gauge-field theory of the YM type (see Weinstein 1999), there is no question, as I shall explain now, that both rest on the same general idea. In this section I explain how GR radically changed our general conception of force-fields and in what sense the gauge-field program is a product of this revolution. It may the case that gravity can not longer be seen as a force-field (Maudlin 1996), but do gauge fields of the YM type have to be force-fields? What exactly makes something a force-field?[8]

      The ontological picture of the pre-relativistic physics is simple: matter exists and evolves in a (flat) background of space and time with interactions among them being mediated by force-fields. It is mathematically embodied in the Lagrangian (or the Hamiltonian) formulation . The action -- which is the time integral of the Lagrangian or, in the case of fields, the space integral of the Lagrangian density -- can always be written as the sum of two terms: $I_{total} = I + I_I$, where I is the action for the dynamics of the given system and I_I the action for the interaction (or coupling) of the system with the force-field. The force-field is characterized by its own action, I_f, which can be added to the others to complete the picture: $I_{total} = I + I_I + I_f$. Applying the variational principle to it, i.e. $\delta I_{total} = 0$, one can obtain the field equation and the equation of motion. (This scheme remains essentially the same in SR, where the background is a 4-dimensional Minkowski spacetime.) With the advent of GR, this scheme had to be radically revised.

      Because of the identification of gravitational effects with the geometrical properties of a 4-dimensional pseudo-Riemannian spacetime, it no longer makes sense to separate the action for dynamics, I, from the one for the interaction with gravity, I_I, since gravity is no longer conceived of as affecting motion as an external force in spacetime. We now have the action of matter in gravity alone as $I_{total} = I_g + I_m$, where I_g is the field action and I_m the matter action. δI_m in terms of δg (for gravity) and δp (for motion) produces two independent terms, i.e. $\delta I_m = \int (C\delta g + D\delta p)d^4x = 0$. Term 'C' becomes the mass-

energy tensor in Einstein's field equation, and term $'D\delta p'$, with some variable

substitutions, yields the geodesic equation (i.e. the equation of motion) (cf. Dirac 1957 and

Wald 1984, 450ff). Beyond this point, the traditional scheme mentioned above seems to

sneak in from the back. When charged matter is present, we have

$I_{total} = I_g + I_m + I_{em} + I_q$, where $\delta I_{em} = 0$ gives one the electromagnetic field equation in a

chargeless region and a mass-energy contribution to the Einstein field equation, and

$\delta I_q = \int (E\delta A + F\delta p)d^4x = 0$ analogously yields a charge current contribution (by the term

'E') to the electromagnetic field equation and the Lorentz force term (via $'F\delta p'$) to the

equation of motion (cf. Dirac 1975, pp. 50-57). This last term is the interaction term that

accounts for the deviation of material objects from their geodesics. A natural extrapolation

of this scheme is that the total action of any number of different charged matter can be

written as $I_{total} = (I_g + I_m) + (I_f + I_q)$, where I_f is the total action of all *other* force-fields

and I_q the total action of all charged matter.[9] Each of these fields contributes to gravity as

the added stress-energy, and it modifies the dynamics of material bodies (deviation from

their geodesics) as external force-fields.

We witness here an interesting twist in the history of physics, where a perfect

unification between inertia and gravity in GR made it more difficult, if not impossible, for a

broader and more important unification -- the unification of all fundamental forces in

nature. When the traditional scheme of action principle is replaced by the GR scheme,

gravity drops out of the list of external force-fields. Instead of the ontological picture

implied by the scheme:

[Spacetime (inertial)] + [Dynamics] + [Interaction] + [Force-fields],

we now have

[Gravitational field] + [Matter] + ['other' Force-fields],

where spacetime drops out; or equivalently:

[Spacetime] + [Matter] + [(3) Force-fields],

where gravity disappears.[10] There is nothing inconsistent about either version of the new ontological picture, but they break a symmetry among force-fields that seem so natural in the traditional picture, namely, the dynamics of a physical system is the result of interactions among material objects via the mediation of force-fields. In this sense, gravity seems to be no different from any other force-field. A demand of unification in sense [III] above (and by the example of relativistic unification of electromagnetism) seems to be legitimized by such a symmetry. On the other hand, this intuition of the symmetry may be ill-founded. When one reflects on the fact that the charge of gravity (i.e. the gravitational mass) necessarily equals the inertial mass, one should realize that gravity is radically different. In Hoefer's words, 'this [the universality of gravity] may be Nature's way of knocking our heads and telling us that gravity is something entirely different from the other forces.[11]' Maudlin (1996), as I quoted him earlier, holds the same view. But many physicists, for various reasons, have refused to accept the exclusion of gravity. Ideal (iii) given above would be a perfect solution of this dilemma, for if (iii) is realized, our ontology then becomes a simple and elegant one:

$$[Spacetime] + [Matter].$$

.

4. The gauge-field program. The above ideal of course has not been, and is unlikely to be, realized, but the gauge-field program which seems at present to have the best hope of unifying the 'other' force-fields is, in spirit though not in content, pursuing the same ideal. Just as in a gauge-field theory where the force arises from the local gauge invariance of the laws of physics, gravitational effect arises from the local Poincaré invariance of the laws as well. I shall explain in detail the latter in the next section. Here I briefly sketch the gauge-field program, which produces theories and unifications for gauge fields of the YM type.

We begin with a picture of the physical world as having particles as quantum fields existing in a 4-dimensional Minkowski spacetime, M. The dynamics of these matter-fields

are determined if the Lagrangian of each type is given, a scheme virtually the same as the classical one discussed in the previous section, except here we Lagrangian densities of fields. Beside being invariant under the global Poincaré group of transformations, a requirement for any relativistic field theory, the Lagrangians also enjoy symmetries of other kinds. For instance, electrons enjoy the symmetry of the U(1) group of global transformations: $\psi(x) \rightarrow \gamma\psi(x)$, $\gamma = e^{-i\Lambda}$, where Λ is an arbitrary constant, and leptons enjoy the symmetry of the SU(2) group of global transformations of similar form, and so on; more on this later. These global symmetries produce conservations of the corresponding charges. Once the constant factors in those transformations are made functions of x ∈ M, an act commonly referred to as *'localizing'* or *'gauging,'* new degrees of freedom open up. The aim of the gauge-field program is to geometricize the 'other' force-fields in such dimensions, the general scheme of which runs roughly as follows (for details, see Göckeker & Schücker 1987; Frankel 1997).

Given a spacetime M and a Lie group G[12] into which M is mapped by local gauge transformations $\gamma \in G^M$, a connection 1-form, A, is introduced (by the demand of gauge invariance) on G which transforms as:

$$A \rightarrow A' = \gamma A \gamma^{-1} + \gamma d\gamma^{-1}. \tag{1}$$

Then, the exterior covariant derivative D is introduced for G as:

$$D\Psi := d\Psi + A \wedge \Psi, \tag{2}$$

Ψ is any vector-valued p-form (i.e. it takes values in a vector space which is a representation of G) and d is the ordinary exterior derivative which maps a p-form to a (p+1)-form and '∧' is the wedge operation. With that, the curvature of G can be simply defined as: F: = DA, a 2-form. In other words,

$$F = DA = dA + A \wedge A = dA + \frac{1}{2}[A, A], \tag{3}$$

where [...,...] is the Poisson bracket. The Lagrangian of a gauge field then has the following general form:

$$(D\Psi)^+ \wedge *(D\Psi) + \frac{1}{q^2} tr(F \wedge *F) - m^2(\Psi^+ \wedge * \Psi) \tag{4}$$

where *q* and *m* are respectively the charge and the mass of the matter field which is represented here by Ψ; '$*$' is the Hodge star operator on forms (analogous to the indices-raising operator for tensors on M), Ψ^+ the Hermitian conjugate of Ψ. The first is the dynamics term, the second the field term, and the third the 'mass' term.

The symmetry group for a particular force-field is determined by the transformational property of the matter-field, i.e., of Ψ; for instance, for electromagnetism, it is U(1) because the charged particles are represented by a Ψ of complex singlet. For the weak interaction, the group is SU(2) because the Ψ of the leptons are complex doublets, and similarly we have SU(3) for gluons of the strong interaction because the Ψ of quarks are complex triplets. The force-fields corresponding to these groups also correspond uniquely to different charges, while the particles may carry one or more charges. With G determined, gauge invariance under G makes it *possible* for the introduction of a connection, which in turn determines the exterior covariant derivative and the curvature. And with these the total Lagrangian, including the dynamics and the force-fields, is determined.

One should note that in the gauge-field program a gauge invariance under G only provides the *possibility* of introducing a connection but by no means determines its form (see Cat 1993; Teller 1999; Liu 1999). This point can best be seen in contrast with the analogous case in GR. The geodesic equation for a free particle in Minkowski spacetime, M, reads simply as: $d^2 u^\sigma / ds^2 = 0$ (where u^σ describes the particle's state in a Cartesian coordinate system, and ds is the proper time), and it is invariant under any rigid (i.e. global) Poincaré transformations. Let the coefficients of the transformations become functions of x $\in$ M, and in so doing we switch to arbitrary coordinate systems. Now, u^σ in an curvilinear system x^μ becomes a function of x^μ. By a simple application of the chain rule, we obtain the following geodesic equation:

$$0 = \frac{d}{ds}\left(\frac{\partial u^\sigma}{\partial x^\mu}\frac{\partial x^\mu}{ds}\right) = \frac{\partial u^\sigma}{\partial x^\mu}\frac{d^2 x^\mu}{ds^2} + \frac{\partial^2 u^\sigma}{\partial x^\mu \partial x^\nu}\frac{\partial x^\mu}{ds}\frac{\partial x^\nu}{ds},$$

which, by being multiplied by a factor $\partial x^\lambda / \partial u^\sigma$, turns into the geodesic equation in GR,

$$0 = \frac{d^2 x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu} \frac{\partial x^\mu}{ds} \frac{\partial x^\nu}{ds}$$

where $\Gamma^\lambda_{\mu\nu}$ turns out to be the needed affine connection for the description of gravity. One may conclude that the 'gauging' of the transformations induces and determines the connection for gravity. The connection for a gauge field cannot be so introduced. Take the U(1) case in which the Lagrangian density of a charged scalar matter field has the form: $(\partial_\mu \psi \partial^\mu \psi - m^2 \psi^+ \psi)$ in some suitable basis, where $\partial_\mu = \partial / \partial^\mu x$. It is invariant under the global U(1) transformations, but when they become local, as functions of x^μ, the chain rule of differentiation does not introduce the connection. To see this, suppose we have $\psi(x) \to \gamma\psi(x)$, $\gamma = e^{-i\Lambda(x)}$, then

$$\partial_\mu \psi \to \partial_\mu \psi' = \partial_\mu (e^{-ie\Lambda(x)}\psi) = (\partial_\mu \psi - ie\psi\partial_\mu\Lambda)e^{-ie\Lambda(x)},$$

and hence,

$$\partial_\mu \psi \partial^\mu \psi \to (\partial_\mu - ie\partial_\mu\Lambda)\psi * (\partial^\mu - ie\partial^\mu\Lambda)\psi.$$

However much the rhs of the arrow resembles the result of the 'minimal coupling,' i.e. $\partial_\mu - ieA_\mu$, it is not. If we define the connection as $A^\mu = \partial^\mu\Lambda$ (i.e. $A = d\Lambda$ in the abstract formulation), what is introduced is a identically flat connection because the curvature in this case is $F = dA = d^2\Lambda \equiv 0$.

The quantum weak force-field[13] operating among leptons and quarks is considered to be unified with quantum electrodynamics into a theory of electroweak force-field under the structure group, SU(2) ⊗ U(1). What kind of unification is it, if it is one? Since its structure group is a Cartesian product of two groups, its exterior covariant derivative is the sum of the two separate minimal couplings, i.e. from (2),

$$D\Psi = d\Psi + (A_e \wedge \Psi) + (A_w \wedge \Psi),$$

and its Lagrangian is also a sum, i.e. from (4),

$$(D\psi)^+ \wedge *(D\psi) + \frac{1}{e^2} tr(F_e \wedge * F_e) + \frac{1}{q^2} tr(F_w \wedge * F_w) - V(\psi^+ \wedge * \psi),$$

where $F_e = DA_e$, $F_w = DA_w$, V(...) is a quadratic function of $\psi^+\psi$, and $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ is a

complex doublet. For the U(1) field of electromagnetism, from (1) above we have

$A_e = ieA$, where e is the charge and A the potential vector of the electromagnetic field; and

$F_e = dA$, since the U(1) field is Abelian, $A \wedge A = 0$. For SU(2) field of the weak force,

we have $A_w = iqW^k\tau_k$, where q is the charge of the weak source, W^k and τ_k (k = 1, 2, 3),

are respectively the potentials (a triplet) and the generators of the weak field; and

$F_w = dA_w + (1/2)q[A_w, A_w]$[14].

The SU(2) part of the electroweak field is the typical YM gauge theory for the weak

isotopic spin space, where when electric charge is ignored there is a perfect local symmetry

between the particles in a doublet, such as $\{e, v\}$, the electron and its neutrino. Besides the

phenomenological charged weak bosons, $W^\pm$, where $W^\pm := (1/\sqrt{2})(W^1 \pm iW^2)$, which

were known before the theory was invented, the theory predicts a neutral weak boson, W^3,

which is, moreover, *mixed* with the photon. One finds the mixing in the neutral interaction

terms of the Lagrangian density, such as the term for a pair of neutrino and antineutrino.

There, we have the phenomenological (or observable) neutral weak field as

$$Z^0 = (eA + qW^3)/\sqrt{q^2 + e^2} ; \qquad (5)$$

and then, the phenomenological photon -- which must be 'orthogonal' to Z^0 -- comes out

naturally as

$$A^{em} = (qA - eW^3)/\sqrt{q^2 + e^2} . \qquad (6)$$

Accordingly, the phenomenological electric charge and weak charge are different from q

and e: the electric charge is $E = qe/\sqrt{e^2 + q^2}$, and the neutral weak charges are

$G_v = \sqrt{e^2 + q^2}$ (coupling with neutrinos), $G_{e_R} = e^2/\sqrt{e^2 + q^2}$ (with right-handed

electrons), and $G_{e_L} = (e^2 - q^2)/\sqrt{e^2 + q^2}$ (with left-handed electrons). (The difference of

the last two is the result of the parity violation in weak interactions.)

To complete the picture, one needs to add the Higgs mechanism which

spontaneously 'breaks' the gauge symmetry of the original theory and therefore makes it

possible for the W-bosons to have non-zero masses, the details of which should not concern us here (cf. Higgs 1974; Aitchison 1982; Quigg 1983). Let us return to our question of whether the electroweak theory is a unification.

There is a sense in which it is a unification of type [I] given above. There seem to be two different levels, one phenomenological at which we have the four bosons, $W^{\pm}$, Z^0, and A^{em}, and one theoretical at which we have the A-field and W^k-fields, where the neutral boson and the photon are, from (5) and (6), *prima facie* composed of A and W^3. However, this picture may not hold up for two reasons. First, it is not clear whether or not the A- and W^k- fields are real. This is similar to a question one asks in the classical electrodynamics, namely, whether the electromagnetic potential A (quite analogous to the A-field here) is real given what is observable there are the strength of **E** and **B**.[15] Without having a handle on the question, there would not a point to ask whether a theory of electrodynamics in terms of **E** and **B** is reducible to a theory of it in terms of A (and thus becoming unified). Second, even if the A- and W^k- fields are real, there is no ontological attribution in this case: four bosons are reduced to another set of four bosons. Hence, even if there is reduction, there is no unification.

Sense [IV] of unification does not seem to apply in this case because even before the electroweak theory, the gauge theories of the separate fields are already well unified in sense [IV]. And it is certainly a unification of type [II], but is it strong enough to be a unification of type [III]. If it is a type [II], it must be because of the mixing roughly characterized above, which is accentuated by the use of the Weinberg (or the weak) angle: $\theta_w = e/q$. From (5) and (6) we see that the mixing of A-field and the W^3-field can be seen as the effect of a rotation of the Weinberg angle. One may think that for this reason, the unification is as good as the one for electromagnetism in SR.[16] This is not the case. What makes the SR unification of electromagnetism profound is the idea that in any situation where an electromagnetic force-field is present, how much of the force is magnetic and how much electric depends entirely on the reference frame in which the force is

observed. Therefore, the electromagnetic force-field is not just a *mixture* of the electric and the magnetic, but really one force-field that manifests itself differently in different reference frames. The electroweak theory, however, portrays a genuine mixture of two different force-fields with two different coupling constants and ranges of application. The mixing -- i.e. the angle -- is also fixed the charges of the two fields: e and q.

For lack of space, I shall not discuss the strong interaction and its unification with the others. Suffices it to mention that the SU(3) theory for the strong interaction is also of YM type and the SU(3) $\otimes$ SU(2) $\otimes$ U(1) unification is very much like the SU(2) $\otimes$ U(1) unification.[17] To summarize, the gauge-field program offers a *two-step* procedure for the unification of force-fields: first we find out whether or not some fields are gauge fields of the YM type (whose general scheme I sketched above), and then if they are, we find out whether or not we can formulate a single theory for them whose gauge group is either a single group or a Cartesian product of their own gauge groups. Note that all YM type fields (i.e. the products of the first step) may be regarded as already unified in sense [IV], but the real meaning of step one is not for unification; it is rather to provide a 'unified' answer to the question: what is it to be a force-field? I may even venture to say that it answers the question: whence does a force-field come? For the answer from the gauge-field program seems to be that it comes from *gauging*: the production of extra dynamical degrees of freedom via localizing appropriate gauge symmetries. However, as I explained above, gauging does not determine uniquely the connection but only makes it possible for its introduction, and hence gauging of the appropriate dynamical degrees of freedom is not sufficient for a gauge theory of the corresponding field. It does, however, appear to be necessary, i.e. a gauge field cannot be introduced without a local gauge symmetry. Therefore, gauging (or local gauge symmetry) is indeed essential to the presence of a force-field. More on this in section **6**.

5. Gauge theories of gravity. Can gravity be unified with the gauge fields in the program? To answer it we need to first find out whether it can be made into a gauge field as the others in the program. Utiyama (1956) is widely regarded as the first person to have explicitly proposed a gauge theory of gravity. His theory was later improved by Kibble (1961) and Sciama (1962). Trautman later argued that the Utiyama-Sciama-Kibble theory can be recognized 'as being essentially equivalent to Cartan's theory of 1923.' (Trautman 1980, p. 305). The Cartan theory -- now known as the Einstein-Cartan theory (see also Trautman 1972, 1973; Göckeler & Schücker 1987; Prasanna 1989) -- is a theory of gravity on a Riemann-Cartan spacetime, U^4, which is a paracompact, Hausdorff, connected, C^∞, and oriented 4-dimensional manifold on which are defined a local Lorentz metric g and a linear affine connection Γ, which transformed just like A in (1), with $\gamma \in GL^+(4, R)$ -- the set of all arbitrary linear transformations on the manifold with positive determinants and which is *metric*, i.e. $Dg = 0$. From Γ a curvature and a torsion tensor can be obtained on U^4. If the torsion tensor vanishes, the Riemann-Cartan spacetime becomes the pseudo-Riemannian spacetime, V^4; if the curvature tensor vanishes, it becomes the Weitzenböck spacetime, A^4; and if both tensors vanish, it becomes the Minkowski spacetime, M^4 (see Hayashi & Shirafuji 1979; Ivanenko & Sardanashvily 1983). The pseudo-Riemannian spacetime, V^4 (i.e. with a signature of (-, +, +, +)), is the one used by GR to describe, inter alia, gravity.

For more details, let us begin with U^4 and its local tangent space, $T_x(U^4)$. A frame can be define on it by a linear base, $\{e_i\}$ (i = 0, 1, 2, 3). With the Lorentz metric tensor, g, we can also define a *coframe* for the cotangent space at $x \in U^4$ with a base, $\{\theta^i\}$, such that $\theta^i = g^{ij}e_j$. In local coordinate system x^μ, we have $\theta^i = h^i_\mu dx^\mu$, where the components, h^i_μ, are known as the tetrads or vierbeins. The two systems of indices -- Latin and Greek -- denote two different sets of transformational properties. The Latin indices are used to indicate tensors in the local coframe which transform as the *G-tensors* (where 'G' indicates the gauge group), and we use the Greek indices to denote tensors

which transform as *covariant tensors* on U^4 under $Diff(R^4)$ -- the set of all diffeomorphisms. Given the group, which may be $GL^+(4, R)$ or one of its subgroups, a connection Γ may be determined; and in U^4 Γ is independent of the metric. Then the curvature is similarly, see (3), defined as $R = d\Gamma + \frac{1}{2}[\Gamma, \Gamma]$ and $R' = \gamma R \gamma^{-1}$, and the torsion is defined as a 2-form with values in R^4, $T := D\theta = d\theta + \Gamma \wedge \theta$ and $T' = \gamma T$. In words, the torsion of a space is represented in the twisting of its local coframe. The field action for the Einstein-Cartan field can be written as

$$I_{EC} = -(1/32\pi G)\int_S tr[R \wedge *(\theta \wedge \theta)],$$

where G is the gravitational constant. The appearance of two different 2-forms is because independent representation of the connection and of the metric are needed, and R is a 2-form in terms of Γ and $\theta \wedge \theta$ a 2-form in terms of g. Now, $\delta I_{EC} = 0$ under the variation of θ alone gets us Einstein's field equation, while $\delta I_{EC} = 0$ under the variation of Γ alone gets us torsion equation.

Different gauge groups, G, when applicable to the Riemann-Cartan manifold may produce slightly different versions of the Einstein-Cartan theory, but there is no gauge group (cf. Invanenko & Sardanashvily 1983) which directly produces the connection in GR (i.e. the Christoffel symbols) and, hence, GR itself. This is rather different from the YM type gauge fields. Let us now see briefly what these different versions are.

The gauge group Utiyama (1956) used to generate the Riemann-Cartan spacetime is that of homogeneous Lorentz transformations (LTs), sometimes known as the Lorentz group[18]. Here, the tetrads comprise the local *orthonormal* coframe in which the transformations among bases are restricted to the 6-parameter LTs. The coordinates in each local frame -- where $g = \eta = (-1, 1, 1, 1)$ -- are related to the global curvilinear coordinates by the vierbein fields, $\theta^i = h^i_\mu dx^\mu$. The use of the Lorentz group is natural if one takes literally the idea of 'gauging' -- i.e. the localization of a previously global symmetry. If the gauge theory of electrodynamics is the exemplar to be followed strictly, GR should be the result of a simple gauging of the global LTs in SR. However, it was argued that the

vierbeins are put in 'by hand', because they are introduced by demanding that tensors in the local coframes be covariant under arbitrary diffeomorphisms across spacetime; and this is clearly a demand beyond the idea of gauging. There is no analogous move of this kind in the YM type gauge fields.

Kibble (1962) and Sciama (1962) noticed the defect and discovered a remedy in the 10-parameter group of inhomogeneous LTs (the Poincaré group). The inhomogeneous LTs, written as infinitesimal transformations in Cartesian coordinates, have the form: $x^i \rightarrow x^i + \Lambda^i_j x^j + \varepsilon^i$, where Λ^i_j is the rotational part and ε^i the translational part. Gauging the group means that the two parts become arbitrary functions of x. The result, to put it very roughly[19], is essentially a manifold which admits local coframes as copies of the Lorentz group, i.e. the rotational (or homogeneous) part, but which also admits general coordinate transformations, i.e. the translational part. 'On the other hand,' Trautman (1973) pointed out, 'curvature and torsion are related, respectively, to the groups of homogeneous transformations and of translations in the tangent spaces of a manifold endowed with a linear connection. (p.194)' Once the local coframes are naturally related to the base via the vierbeins, $\theta^i = h^i_\mu dx^\mu$, the rest of the geometry of U^4 is derived as in Utiyama.

The Poincaré group approach was further developed by Hehl et al (1976). In that work an active (rather than a passive) interpretation of the localized Poincaré transformations is used to derive the geometry of the Riemann-Cartan spacetime. The general idea of gauge-field program -- the invariance of physics under localized transformations -- produces in this case two gauge potentials: the rotational potential for the connection and a translational potential for the tetrads. Hehl et al further argue that the contortion (because of the torsion in U^4) is a physically meaningful magnitude; it couples to the spin of matter just as the curvature couples to the mass. We do not see the contortion part in GR because GR is a theory of the macroscopic, and at the macroscopic level the total spin of matter is usually unnoticeable (e.g. being averaged out).

One may think that this is the end of the story for gauge gravity; what remains is the experimental verification (or falsification) of the coupling between spin and contortion. No, it is not, for as early as 1928, Einstein entertained the idea of characterizing gravity in absolute parallelism, i.e. in a spacetime with a vanishing curvature (see Pais 1982, pp. 344-345).[20] The idea was later picked up and developed into a full-fledged gauge theory of gravity (cf. Hayashi & Shirafuji 1979; Muench et al 1998). It turns out -- details aside -- that the Weitzenböck spacetime has enough geometry of the right kind to describe gravity. This result may appear surprising because gravity has always been associated with the Riemannian curvature; but here the curvature identically vanishes. However, if one begins with a Riemann-Cartan spacetime -- whose connection has no intrinsic relation with the metric, there is no *a priori* reason that gravity be captured only by a curved spacetime. The so called *teleparallel* tetrad fields can do the job equally well, according to Hayashi and Shirafuji (1979). One must also note that in a Weitzenböck spacetime, the vanishing of the Riemannian curvature does not at all imply that spacetime is flat (i.e. Minkowskian). The metric is defined by the tetrad fields, as is the case for all Riemann-Cartan manifolds. One may say that in the teleparallel version of gauge gravity, it is the '*twisted*' rather than the '*curved*' manifold that accounts for gravity.

Along the same line, an argument is made recently by de Andrade and Pereira (1997, 1998) for the conclusion that gravity can be equivalently described either by the Riemann spacetime or by the Weitzenböck spacetime. In other words, one can use either the Riemann connection (torsion free) or the Cartan connection (curvature free) to describe the effects of gravity.[21]

Some other groups are also considered as candidates for the gauge group, e.g. de Sitter group or the general linear group, $GL(4, R)$, or the general affine group, $GA(4, R)$ (see Trautman 1980b; Ivanov & Niederle 1982; Ivanenko & Sardanashvily 1983; Nounou 1995). Unlike the case of Poincaré group, which is intended to be a replacement of the Lorentz group, the $GL(4, R)$ and $GA(4, R)$ are generalizations. The local coframes with

such groups as their gauge groups are more general than the Riemann-Cartan spacetime; and a contraction of $GL(4, R)$ or $GA(4, R)$ to either the Lorentz or the Poincaré group will have to be implemented before we get the Riemann-Cartan spacetime back.

There are also proposals (cf. Ivanov & Niederle 1982 and Mukunda 1989) in which attempts are made to treat the Poincaré group as an internal symmetry group when it is gauged -- i.e. the elements of the group do not act on spacetime points at all. In other words, the full invariance group of the gauge gravity Lagrangian is the product of two groups: $P^{loc} \otimes Diff(R^4)$, where P^{loc} is the gauged Poincaré group and $Diff(R^4)$ the set of all diffeomorphisms acting on the spacetime manifold. The point of doing this is to make the result of gauging in gravity having exactly the same structure as that in the YM type gauge fields. In the later case, we have, $G^{loc} \otimes P$, where G^{loc} is the gauge group of internal symmetries and P is the global Poincaré group which acts on the Minkowski spacetime. I can see several problems with this approach, the most obvious of which would be this: what would be the group structure for an interaction between, say, electromagnetism and gravity? Should it not be $U(1) \otimes Diff(R^4)$, according to the approach? But according to the same approach, gravity disappears in the Lagrangian under this group.

There is yet another very different approach to making GR a gauge-field theory, which is carried out on Riem(Σ), the space of Riemannian metrics on a compact 3-dimensional manifold. Since this approach is, at least in appearance, quite different from the gauge-field program I examine in this paper, I shall refer the interested readers to the following references: Earman 2000; Belot & Earman 1999; Wald 1984; and the literature therein. It appears that this approach disregards the symmetry of the Poincaré group from the very beginning by making an explicit separation of the temporal dimension from the spatial ones; but see Crnkovic & Witten 1987 for an argument to the contrary.

6. The nature of gauging and the unity of force-fields. We now return to the question: can gravity be unified with the other fundamental force-fields in the gauge-field program? An affirmative answer is supposed to be given by gauge gravity, at least with regard to the first step of the unification in the program (see the end of section **4**). The Einstein-Cartan gauge theory of gravity may not be true, but that is a possibility for all gauge-field theories in the program. Hence, our question should be: if the Einstein-Cartan theory is true, is it a gauge theory that truly belongs to the gauge-field program?

Two essential features of the YM type gauge fields are pertinent to our evaluation of gauge gravity.

[F1]   The background spacetime for such fields may be Minkowskian, i.e. flat. (They are also compatible with curved background spacetime, such as the pseudo-Riemannian spacetime, but that is not important here.)

[F2]   The existence of such fields is made possible by the gauging of the appropriate gauge spaces.

Gauge gravity cannot have the first of these features because if one 'flattens' the background spacetime for the Einstein-Cartan theory, none of the geometric objects on the coframes can have the proper properties to describe the effects of gravity. Since the coframes with a Minkowski metric will be $\theta^i = h^i_\mu dx^\mu = \delta^i_\mu dx^\mu = dx^\mu$, i.e. they coincide with the global Cartesian frame (which is holonomic and orthonormal), both the curvature and the torsion will vanish.[22]

Perhaps this fact gives some explanation to the real difference between the *external* and the *internal* symmetries which were said to separate GR from the other gauge-field theories. While the other gauge fields are geometric properties of spaces over and beyond spacetime -- hence, internal to the matter they affect -- gravity cannot be conceived as being independent of spacetime -- hence, external to the matter it affects. And while for gauge fields it is still possible to retain the sense of 'force' as something that compel objects to

deviate from their inertial motion in a spacetime, i.e. from their geodesics, that is not at all possible for gravity.

Does the gauge gravity have the second feature? From the works we discussed above, the answer seems to be 'Yes.' The question is: which group should it be, the gauging of which 'automatically' yields the Einstein-Cartan theory of gravity? We have seen that it cannot be the Lorentz group, for its gauging in the coframe of a flat spacetime is compatible with the spacetime's remaining flat; but no gravity arises this way. As mentioned earlier, if one gauges the Poincaré, rather than the Lorentz, group, this defect can be removed. We now examine some details of this argument to see how it is done. To my best knowledge, Hehl et al (1976) and Mukunda (1989) give the most detailed arguments. Hehl et al begin their argument by claiming,

> We shall show in this section that to require the invariance of a special relativistic theory of matter under local space-time rotations and translations (independent actions of the Poincaré group at every point) leads *inexorably* to the introduction of torsion and curvature. In view of the fundamental significance of the Poincaré group for physics, and of the local gauge approach in general, we take such a derivation as the strongest evidence that the geometry of the physical world is indeed a Riemann-Cartan geometry. [my italics] (p.400)

To show this, they begin with the global inhomogeneous LTs of the Poincaré group $x^i \rightarrow x'^i = \Lambda^i_j x^j + a^i$, where Λ^i_j are the rotational, and a^i the translational, elements of the group. Making these infinitesimal transformations, we have, $\delta x^i = \omega^i_j x^j + \varepsilon^i$; and for a scalar matter-field, $\psi(x)$, we have

$$\psi(x) \rightarrow \psi'(x) = (1 + \omega^{ij} f_{ij}) \psi(x), \qquad (7)$$

where f_{ij} are the generators of the group. There are two ways of gauging the Poincaré group. If one treats the LTs as transformations among coordinate systems (i.e. the passive

reading), then making the transformations local would blur the distinction between the rotational and the translational elements. It turns the group into a 4-parameter group of general coordinate transformations: $x^\mu \to x'^\mu = x^\mu + \varepsilon^\mu(x)$, where the Greek indices signify coordinates in arbitrary coordinate systems.

However, if one interprets the infinitesimal LTs as acts of re-orienting and displacing our experimental (or measuring) frames or apparati (i.e. the active reading) and *keeps the two actions separate,* one first writes (7) as,

$$\psi(x) \to \Pi\psi(x) = (1 + \omega^{\mu\nu}f_{\mu\nu} - \varepsilon^\mu \partial_\mu)\psi(x), \qquad (8)$$

where Π is a linear mapping of the field at the neighborhood of x, which *re-orients* and *shifts* $\psi(x)$ with the following respective amounts:

$$\omega^{\mu\nu} = \omega^{ij} \qquad \text{and} \qquad \varepsilon^\mu = \varepsilon^i + \delta_i^\mu \omega_j^i x^j, \qquad (9)$$

where ε^i and ω^{ij} (and ω_j^i) are the same parameters as in (7), which are now restricted to an orthonormal basis, $\mathbf{e}_i$, at x. This operation is explained by Hehl et al (1976) thus:

Replace the fields ψ at a point x^i by fields which have first been rotated by an amount $-\omega$, that is, $\psi(x) \to [\Lambda\psi](x) = (1 + \omega f)\psi(x)$, and then have been translated by an amount $+\varepsilon = \overset{\circ}{\varepsilon} + \overset{\circ}{\omega} \bullet x$, that is, $[\Lambda\psi](x) \to [\Pi\psi](x) = [\Lambda\psi](x - \varepsilon)$. Then, experience shows, matter distribution $\Pi\psi$ and ψ are physically equivalent (p. 401).

Because of the equivalence of the active and the passing reading, nothing is changed in SR. However, when one gauges the symmetry in the active reading, one localizes the ε^i and ω^{ij} (and ω_j^i) in the orthonormal frames. As usual, when $\varepsilon^i \to \varepsilon^i(x)$ and $\omega^{ij} \to \omega^{ij}(x)$, (note that they are all G-tensors), we have the Lagrangian density:

$$L(\psi, \partial_i \psi) \to L(\psi, \partial_i \psi, \Gamma_\mu^{ij}, h_\mu^i),$$

where in the limit of a weak force-field, we have,

$$\delta\Gamma_\mu^{ij} \cong -\partial_\mu \omega^{ij}(x) \qquad \text{and} \qquad \delta h_\mu^i \cong \omega_j^i \delta_\mu^j - \partial_\mu \varepsilon^i(x),$$

as the results of non-vanishing differentials of ε^i and ω^{ij}. This is the first step through which the Poincaré group is said to have introduced the connection coefficients Γ_μ^{ij} and the vierbeins h_μ^i as coordinate vectors via the localization of ε^i and ω^{ij}. (Again, Latin indices denote tensors in the orthonormal coframes and Greek indices tensors in the general coordinate systems.)

Further advantage is taken of this active reading of the Poincaré group, for (8) and (9) may be seen as yielding a rigidity condition which says that Π changes 'neither the distance between events nor the relative orientation of neighboring matter fields (Hehl et al 1976, p. 402).' More formally, the condition before gauging is:

$$\Pi(\xi^i \partial_i \psi) = (\Pi\xi^i)\partial_i(\Pi\psi) = (\Pi\xi^i)\delta_i^\mu \partial_\mu(\Pi\psi),$$

for any infinitesimal vector field, ξ^i, that is bound to the field ψ. Here the insertion of δ_i^μ in the last step is trivial with the global Poincaré group. However, when the group is gauged, the rigidity condition should hold as well. But then we have

$$\delta_i^\mu \rightarrow h_i^\mu(x) \quad \text{and} \quad \partial_\mu \rightarrow D_\mu = \partial_\mu + \Gamma_\mu^{ij}(x)f_{ij}, \tag{10}$$

such that the rigidity condition holds in general as,

$$\Pi(\xi^\mu D_\mu \psi) \cong (\Pi\xi^i)(h_i^\mu + \delta h_i^\mu)(D_\mu + \delta\Gamma_\mu^{ij}f_{ij})(\Pi\psi).$$

Here the first transition in (10) introduces the vierbeins which connect the orthonormal coframes to the general coordinate base, while the second transition in (10) is the result of a relative rotation, $dx^\mu \Gamma_{\mu j}^i \mathbf{e}_i$, of the frame base, $\mathbf{e}_j$, at neighboring points x^μ and $x^\mu + dx^\mu$. Once the connections and the vierbein fields are introduced, the rest of the steps towards the full Riemann-Cartan manifold are rather standard, which most of the accounts of the Einstein-Cartan theory follow. For instance, the metric of the background spacetime is given by

$$g_{\mu\nu} = \eta_{ij}h_\mu^i h_\nu^j,$$

In other words, the pseudo-Riemannian metric is a natural result of the gauging of the translational element of the Poincaré group.[23] Although Weinstein (1999) is right in arguing that $Diff(R^4)$ is not the gauge group of any fibre bundle that may describe gravity, gauge

gravity is at least able to tell a story of how $Diff(R^4)$ drops out for the background spacetime as a result of gauging the Poincaré group.

Therefore, with respective to F2, gauge gravity -- the Einstein-Cartan theory of gravity -- appears to be just as good a gauge-field theory as any of the YM type; but it differs from them with respect to F1. This I believe clarifies the precise sense in which such claims as gravity in GR is a gauge field but not one of the YM type. The correct claim should be: gravity in the Einstein-Cartan theory is a gauge field (because of having F2) but it is not one of the YM type (because of not having F1). Note that the GR gravity has neither features.[24]

Why should F1 be an essential feature to the gauge-field program (or why is it important that gauge fields are of YM type)? I have ventured a short answer to this earlier, namely, it allows us to retain the traditional sense of force for these gauge fields. Geometricized or not, they can still be understood as agents which cause deviations of quantum particles from their inertial evolutions in spacetime. Such an interpretation is no longer available for gauge gravity because of the modified notion of geodesics in GR (or the Einstein-Cartan theory). This difference is also fundamental for another reason. Gauging is an act of making spacetime independent (global) symmetries into spacetime dependent (local) ones. If the symmetry is of a gauge space other than spacetime, such as the U(1) symmetry of phase space, it make sense to say that the gauging of the symmetry produces a transition from a 'flat' phase space to a 'curved' one in the *same spacetime background*. We cannot use the same image for the gauging of the Poincaré symmetry, since it is a symmetry *of* spacetime *in* spacetime. As I discussed earlier, the gauging that changes a global coframe into local ones, $\theta^i : \delta_\mu^i dx^\mu \rightarrow h_\mu^i dx^\mu$, inevitably 'unflatttens' the metric, because the *soldering* relation between the vierbeins, h_μ^i, and the metric.

We are now ready to revisit the questions I asked at the beginning of section **3**. They are essentially two: what is a force-field, given what we have learned so far? And given the answer to that, is gravity a force-field? In section **3** we see that the advent of Einstein's GR in some sense breaks an old unified concept which takes forces to be external interactions that

alter the inertial state of a system. Though the concept was a unified one, the forces were distinct -- there was no theory which unifies them in the pre-relativistic era. After GR, the quest was not only for a new unified concept but also for a unified theory. If Einstein's dream for a unified field theory were fulfilled, there would be a unified theory for a single 'force'-field in which *the force disappears*. If all fundamental forces were reduced to some geometric properties of spacetime -- i.e. a particle freely moving in any or all force-fields is moving along its geodesic -- it would be better to stop using the notion of force or force-field altogether. Although the picture in the gauge-field program is more complicated, it is still reasonable, I argue, to regard the above line of thought valid in it. The notions of reduction and of geometricization are almost the same as in Einstein's original conception, only the spaces in which the force-fields are geometricized are different. Therefore, I urge that, at least for the YM type force-fields, we see the first stage of the program (see the end of section **4**) as an attempt to *replace* one unified concept of forces with another, perhaps more profound, one.

In view of the gauge-field program, we now realize that the GR reduction involves two separable components, one is the geometrization of force-fields (i.e. gravity) and the other is the identification of the geometry with that of the pseudo-Riemannian spacetime. The sense in which the gauge-field program provides a replacement of the unified concept of forces only involves the first component -- i.e. geometricization. Redhead (1999) gives a general sketch of the nature of such a geometricization in the mathematical modeling of physical systems. According to his account, surplus structures are often inevitable in the process of imbedding a physical system into a mathematical model. Such structures are always 'controlled' by using groups of global transformations such that the contracted model (one without the surplus) becomes isomorphic to the physical model. But such surpluses are fertile grounds for gauge fields, for once some of them are gauged -- i.e. that the transformations are made local -- their geometries may offer the grounding for the reduction of the appropriate force-fields. Because of F1, the gauging of internal symmetries realizes the first component and therefore geometricizes the corresponding force-fields, but it does not quite eliminates them, since it does

not realize the second component. There is still a proper notion of the geodesic of a system, whether in a flat or a curved spacetime, from which the deviation of the system's motion is accountable by the presence of some gauge fields of the YM type.

In the pre-relativistic regime, force-fields are represented either in the interaction Lagrangians, as explained in section **3**, or by potential terms, which express a disposition of making systems under the fields' influence to deviate from their geodesics. The force-fields are not unified in terms of their categorical properties, such as their energy and momentum, but they all play the same *causal role* in nature: they *dispose* objects under their influence to deviate from their geodesics. Geometricization, by which the gauge-field program replaces the old regime, helps to replace the *dispositional* concept of force-fields by a *categorical* one: the shapes of dynamical (or gauge) spaces.

I believe we now know the answers to our questions. They are not straightforward 'yes' or 'no' answers. We know exactly what a gauge force-field is *qua* force and how it differs from the old concept; and we know gravity is strictly speaking not a gauge force-field -- it is not a force-field at all -- but it shares with such fields an essential feature, geometricization.

Finally, from a unificational point of view, the search for a gauge theory of gravity does not seem well motivated. Unless one can find a gauge theory that possesses F1, gravity cannot be regarded as one of the force-fields of the YM type; and therefore there is little hope that it can be unified with the others within the program. Of course, a unification is always possible with field theories beyond the program. But then gauge gravity is even less well-motivated, for one should probably try to unify a gauge field with gravity in GR. At least, GR is a much better confirmed theory of gravity than the Einstein-Cartan theory.

References

Aitchison, I. J. R. (1982). <u>An Informal Introduction to Gauge Field Theories</u>. Cambridge, Cambridge University Press.

Belot, G. and J. Earman (1999). "Pre-Socratic Quantum Gravity." forthcoming in Callender, C. and H. Huggett, eds. Philosophy at the Planck Length. Cambridge, Cambridge University Press.

Callender, C. and N. Huggett (1999). "Why Quantize the Gravitational Field (Or Any Other Field for That Matter)?" PSA 2000 paper (preprint).

Cat, J. (1993). "A Philosophical Introduction to Gauge Symmetries." Lectures delivered at the London School of Economics, October 1993 (preprint).

Choquet-Bruhat, Y., C. DeWitt-Morette, et al. (1982). Analysis, Manifolds and Physics. Amsterdam, North-Holland.

Corry, L. (1999). "From Mie's Electromagnetic Theory of Matter to Hilbert's Unified Foundations of Physics." Studies in History and Philosophy of Modern Physics **30B**: 159-183.

Crnkovic, C. and E. Witten (1987). Covariant Description of Canonical Formalism in Geometrical Theories. 300 Years of Gravitation Eds. S. Hawking and I. Israel. Cambridge, Cambridge University Press, pp. 676-684.

de Andrade, V. C. and J. G. Pereira (1997). "Gravitational Lorentz Force and the Description of the Gravitational Interaction." Physical Review D **56**: 4689-4695.

de Andrade, V. C. and J. G. Pereira (1998). "Riemannian and Teleparallel Descriptions of the Scalar Field Gravitational Interaction." General Relativity and Gravitation **30**: 263-273.

Deser, S. (1970). "Self-Interaction and Gauge Invariance." General Relativity and Gravitation **1**: 9-18.

Dirac, P. A. M. (1975). General Theory of Relativity. New York, Wiley.

Earman, J. (2000). "Gauge Matters." manuscript for the Symposium on the Concept of Gauge in Modern Physics at the PSA 2000 Biennial Meeting, Vancouver, Canada.

Frankel, T. (1997). The Geometry of Physics: An Introduction. Cambridge, Cambridge University Press.

Healey, R. (1997). "Nonlocality and the Aharonov-Bohm Effect." Philosophy of Science **64**: 18-41.

Goldberg, S. (1976). "Max Planck's Philosophy of Nature and His Elaboration of the Special Theory of Relativity," in Historical Studies in the Physical Sciences. **7**R, 125.

Gupta, S.N. (1954). "Gravitation and Electromagnetism." Physical Reviews **96**: 1638-1685.

Hayashi, K. and T. Shirafuji (1979). "New General Relativity." Physical Review D **19**: 3524-3553.

Hehl, F. W., P. von der Heyde, G.D. Kerlick, and J.M. Nester (1976). "General Relativity with Spin and Torsion: Foundations and Prospects." Reviews of Modern Physics **48**: 292-416.

Higgs, P. W. (1974). Spontaneous Symmetry Breaking. <u>Phenomenology of Particles at High Energies</u> Eds. R. L. Grawford and R. Jennings. New York, Academic. 529.

Hilbert, D. (1924). "Die Grundlagen der Physik." <u>Mathem. Annalen</u> **92**: 1-32; reprinted in Hilbert, D. <u>Gesammelte Abhandlungen</u> Bd III, Bronx, NY, Chelsea Publishing, pp. 258-289.

Ivanenko, D. and G. Sardanashvily (1983). "The Gauge Treatment of Gravity." <u>Physics Reports</u> **94**: 1-45.

Ivanov, E. A. and J. Niederle (1982). "Gauge Formulation of Gravitation Theories I: The Poincare, de Sitter, and Conformal Cases." <u>Physical Review D</u> **25**: 976-987.

Kibble, T. W. B. (1961). "Lorentz Invariance and the Gravitational Field." <u>Journal of Mathematical Physics</u> **2**: 212-221.

Kosso, P. (1999). "Symmetry Arguments in Physics." <u>Studies in History and Philosophy of Science</u> **30A**: 479-492.

Leeds, S. (1999). "Gauges: Aharonov, Bohm, Yang, Healey." <u>Philosophy of Science</u> **66**: 606-627.

Liu, C. (1994a). "Is There a Relativistic Thermodynamics: A Case Study of the Meaning of Special Relativity." <u>Studies in History and Philosophy of Science</u>. **25**: 983-1004.

Liu, C. (1994b). "The Aharonov-Bohm Effect and the Reality of Wave Packets." <u>British Journal for the Philosophy of Science</u> **45**: 977-1000.

Liu, C. (1996). "Potential, Propensity and Categorical Realism." <u>Erkenntnis</u> **45**: 45-68.

Liu, C. (1999). "Gravity and Electromagnetism under the Gauge-Field Program." Lecture Notes. Conference on the Philosophical Significance of Gauge Theories in Physics, held at Philosophy Department at the University of Arizona, March 18-21, 1999.

McCormmach, R. (1970). "H.A. Lorentz and the Electromagnetic View of Nature." <u>Isis</u> **61**: 457-497.

Maheshwari, A. (1989). Kaluza-Klein Theories. <u>Gravitation, Gauge Theories and the Early Universe</u> Eds. B. R. Iyer, N. Mukunda and C. V. Vishvershwara. Dordrecht, Kluwer Academic Publishers. 423-447.

Maudlin, T. (1996). "On the Unification of Physics." <u>The Journal of Philosophy</u> **93**: 129-144.

Maudlin, T. (1998). "Discussion: Healey on the Aharonov-Bohm Effect." <u>Philosophy of Science</u> **65**: 361-368.

Muench, U., F. Gronwald, et al. (1998). "A Brief Guide to Variations in Teleparallel Gauge Theories of Gravity and the Kaniel-Itin Model." <u>General Relativity and Gravitation</u> **30**: 933-961.

Mukunda, N. (1989). An Elementary Introduction to the Gauge Theory Approach to Gravity. <u>Gravitation, Gauge Theories and the Early Universe</u> Eds. B. R. Iyer, N. Mukunda and C. V. Vishvershwara. Dordrecht, Kluwer Academic Publishers. 467-479.

Nounou, A. (1995). "Gauge Theories and Gravity." M.A. Thesis (Imperial College, London).

O'Raifeartaigh, L. (1997). The Dawning of Gauge Theory. Princeton, Princeton University Press.

Pais, A. (1982). 'Subtle is the Lord...': The Science and the Life of Albert Einstein. Oxford, Oxford University Press.

Planck, M. (1908a). "Zur Dynamik bewegter Systeme." Annalen der Physik **26**: 1-34.

Planck, M. (1908b). "Die Einheit des physikalischen Weltbildes." in M. Planck, Vorträge und Erinnerungen (1965). Darmstadt, Wissenschaftliche Buchgesellschaft, 18-51.

Prasanna, A. R. (1989). Differential Forms and Einstein-Cartan Theory. Gravitation, Gauge Theories and the Early Universe Eds. B. R. Iyer, N. Mukunda and C. V. Vishvershwara. Dordrecht, Kluwer Academic Publishers. 119-132.

Quigg, C. (1983). Gauge Theories of the Strong, Weak, and Electromagnetic Interactions. Reading, MA, Benjamin.

Rajasekaran, G. (1989). Building up the Standard Gauge Model of High Energy Physics.. Gravitation, Gauge Theories and the Early Universe Eds. B. R. Iyer, N. Mukunda and C. V. Vishvershwara. Dordrecht, Kluwer Academic Publishers. 185-236.

Redhead, M. (1999). "Lecture on Symmetries and Gauge Principles." Lecture notes (preprint)

Sciama, D.W. (1962). "On the Analogy between Charge and Spin in General Relativity." in Recent Development in General Relativity. Oxford, Pergamon + PWN, pp. 425-???.

Teller, P. (2000). "The Gauge Argument." Philosophy of Science **67** (Supplement): S466-S481.

Trautman, A. (1972). "On the Einstein-Cartan Equations. I, II, III" Bulletin de L'Académie Polonaise des Sciences. (série des sciences, math., astr. et phys.) **20**: 185-190, 503-506, 895-896.

Trautman, A. (1973). Theory of Gravitation. The Physicist's Conception of Nature Ed. J. Mehra. Dordrecht, Reidel. 179-198.

Trautman, A. (1980). Fiber Bundles, Gauge Fields, and Gravitation. General Relativity and Gravitation, vol. 1 Ed. A. Held. London, Plenum. 287-308.

Utiyama, R. (1956). "Invariant Theoretical Interpretation of Interaction." Physical Review **101**: 1597-1607, reprinted in O'Raifeartaigh 1997, pp. 213-239.

Vizgin, V. P. (1994). Unified Field Theories (in the first third of the 20th century). Basel, Birkhäuser Verlag.

Wald, R.M. (1984). General Relativity. Chicago, University of Chicago Press.

Weinberg, S. (1996). <u>The Quantum Theory of Fields</u>. Vol. II. Cambridge: Cambridge University Press.

Weyl, H. (1918). "Gravitation and Electricity," English translation in O'Raifeartaigh (1997), 24-37; original in <u>Sitzungsber. Preuss. Akad. Berlin</u> (1918), 465.

Weyl, H. (1919). <u>Space-Time-Matter</u>, English translation, London: Methuen (1922).

Whittaker, E. (1953). <u>A History of the Theory of Aether and Electricity</u>. vol. 2, New York, Humanities Press.

Wigner, E.P. (1967). <u>Symmetries and Reflections</u>. Bloomington: Indiana University Press.

NOTES

† I would like to thank Gordon Belot, Richard Healey, Peter Kosso, Stephen Leeds, Paul Teller for listening to an earlier version of this paper and for their questions and comments.

---

[1] I would like to dedicate this paper to the memory of Wes Salmon. Acknowledgment is withheld for the review process.

[2] I shall consistently use the term 'force-field' (or simply force) in contrast to the term 'matter-field'. In classical (i.e. non-quantum) physics, there is no matter-field, so the simple term 'field' may suffice; but in quantum physics, where we have both matter- and force- fields, the longer term is necessary.

[3] If one finds the claim that gravity is identified with spacetime in GR too crude, here is a better way of stating it. In GR, gravitational effects can be identified with geometrical properties of spacetime regions in a way that the effects of the other three force fields cannot (at least, not likely if the standard model is on the right track). This, I warn, should not be taken as being incompatible with the *metaphysical* claim that all matter and forces are properties of spacetime regions, e.g. a house may be thought of as some kind of property of the space it occupies, but it cannot be said to be a geometrical property of the space unless a theory about that is given.

[4] This predicate is not superfluous here because the claim may not be true for a spacetime of more than 4-dimensions. see the next footnote.

[5] There is a small complication of the otherwise clear dichotomy. If the 'manifold' of nature is a Riemannian manifold of more than 4 dimensions, then it is possible to integrate the other forces into this manifold. One example of such a theory is the Kaluza-Klein theory of gravity and electromagnetism in the 5-dimensional Riemannian manifold. Does this theory, if made viable, count as a successful unification of gravity and electromagnetism? Perhaps it is; but it is certainly not a reduction of the two forces to manifestly 4-dimensional *spacetim***e**. Hence, even if the Kaluza-Klein theory is true, gravity still does not belong to the rest of the natural forces. Because the Kaluza-Klein approach to the unity of force-fields is not one in the gauge-field program, I shall not discuss it in any detail in this paper. See O'Raifeartaigh 1997 and Maheshwari 1989 and the literature therein for details.

[6] Weinstein (1999) acknowledges such attempts, see pp. S152-S154, and points out, 'Though some of these constructions are interesting and illuminating in their own right, in none of them is the diffeomorphism group a gauge group.' I shall address the last point in due course.

[7] I here use 'space/time' to mean the pre-relativistic space and time.

[8] An approach to an answer to this question will be given in section **6**, after we have seen how physicists struggle with this notion. As a starter, the concept of force-field behind the theories I examine in this and the next section is that it is a substance which interacts with matter and, after quantization, also with itself.

[9] This is formulated slightly differently from Dirac's formulation, where he has, $I_{total} = I_g + I'$ (p. 58). Mathematically they are obviously the same, the difference is in emphasis. While Dirac wants to single out gravity, I want to single out gravity and massed terms.

[10] It is pointed out to me that if I consider the Palatini variational principle, I may get a different ontological picture, or at least a more profound understanding of what the theory of GR says. Suppose we only consider I_g. We calculate $\delta I_g = 0$ in terms of δg and $\delta \Gamma$ (i.e. the variation of the connection), separately. The result is that the vanishing of δg's coefficient yields Einstein's field question in vacuum, while the vanishing of $\delta \Gamma$'s coefficient yields the metric compatibility condition, $\nabla_\lambda g^{\mu\nu} = 0$. This may provide us with a better understanding of GR, but I fail to see how it can offer an ontological picture which differs from the two alternatives I provided here. If the metric is interpreted as the potential for gravity, the results of both variations -- the field equation and metric compatibility -- are about gravitational field; but if the metric is interpreted as a property of spacetime, the two results should be understood as about spacetime. I fail to see the possibility in the Palatini method of treating spacetime and gravity as ontologically distinct substances if it is not already in the regular theory of GR. Such a possibility may indeed emerge in the Einstein-Cartan theory, but there it is the result of a different conception of spacetime, as I shall explain in section **5**.

[11] From a lunch-time conversation in October of 1999.

[12] Keep in mind that a Lie group is a special kind of manifold; and in this case, having a copy of G sitting on each x ∈ M (since M is mapped into G by local gauge transformations γ) provides the new degrees of freedom (or new space) to which gauge fields are reduced. For those who are familiar with the fibre bundle language, these are the fibres; and the A and F introduced later are the connection 1-form and the curvature 2-form of the fibre bundle.

[13] In this paper I will not introduce the quantization of these fields in any systematic way. The procedures of quantizing force- and matter- fields have their own deep conceptual problems, but they are not at all pertinent to the question of whether or not the fields are unifiable. What I delineate in the following is of such generality that it is entirely preserved after the quantization of the fields in question. One may therefore think of the fields discussed here as quantum fields, e.g. interactional bosons. For problems in the quantization of gravity, see Belot & Earman 1999.

[14] In a suitable coordinate system the components of this F reads:
$(F_w)_{\mu\nu} = (\partial_\mu W_\nu^k - \partial_\nu W_\mu^k) - q[W_\mu^k, W_\nu^k]$, which is the familiar form one sees in most of the physics textbooks.

[15] For a discussion of the meaning of A, see Liu 1994b, 1996; Healey 1997; Maudlin 1998; Leeds 1999.

[16] Maudlin says, 'It [the electroweak theory] is at least as strong as the unification of electricity and magnetism in Maxwell's theory, ... And there is something deeper, ... Still the unification fails to reach the level of perfection found in GTR,... (Maudlin 1996, p. 138)' So, it is not clear whether he thinks that it is as good a unification as the one in SR -- the other perfect unification.

[17] There have been attempts to reduce the coupling constants to a single one. For electroweak, one may imbed SU(2) ⊗ U(1) into SO(3), but it is 'ruled out by experiment.' (Weinberg 1996) For the strong and the electroweak unification, one may likewise imbed SU(3) ⊗ SU(2) ⊗ U(1) into groups such as SU(5) or SO(10). The jury is still out on such attempts (Weinberg 1996).

---

[18] In Utiyama's paper (1956), the full Riemann-Cartan spacetime is never explicitly derived, nor is torsion even considered. All results are derived under the stipulation that $Dg = 0$ and $\Gamma^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\nu\mu}$. It was later realized that the natural result of gauging the Lorentz group is indeed the Riemann-Cartan spacetime.

[19] Here I mostly follow Trautman 1973. But we will see the details of this point in section **6**.

[20] Einstein appeared to be wanting to see whether absolute parallelism could carry electromagnetic field. But it is not clear how he could do it without curvature on the manifold for gravity. If the torsion tensor was used for electromagnetic field, which Einstein did so use, what could he use for gravity? The metric is still pseudo-Riemannian because $g_{\mu\nu} = h^i_{\mu} h^i_{\nu}$, but it is not clear whether this is sufficient for gravity.

[21] De Andrade and Pereira (1997) give the following interesting argument. First, by entertaining what appears to be a principal bundle (think of it as a bundle of local coframes) whose typical fiber is a local translational group -- i.e. the translational part of the localized Poincaré group -- and the base space a Minkowski spacetime, they are able to derive a 'gravitational analogue of the Lorentz force' (p. 4691) that gives an adequate account of gravity in the YM fashion. Then, the vierbein fields so obtained (as the vector fields that set up the local translational coframes) are used to define, first, a teleparallel covariant derivative and, second, a Riemannian covariant derivative. Both eventually lead to a description of gravity that is then proven to be equivalent. However, as they show, the teleparallel description treats gravity as an external force-field -- external to the Weitzenböck spacetime, while the Riemannian description identifies gravity with the spacetime (as is the case in GR). In other words, if we choose the Weitzenböck formulation, gravity is just like any other YM type gauge fields, whose presence is external to spacetime.

[22] There is a long-standing tradition consisting of attempts to reproduce results in Einstein's GR in a flat spacetime background. It is not at all clear to me whether they can be integrated into the gauge-field program I discuss in this paper. Interested readers may consult Deser 1970 and Gupta 1954.

[23] See also (Mukunda 1989) for a simpler and more intuitive but less rigorous discussion of the gauging of the Poincaré group.

[24] Again, if one consider GR as a theory of constraint Hamiltonian systems on Riem(Σ) (see the end of section **5**), these considerations may not apply. The relationship between this approach and the one I examine in this paper is an interesting one; but I must leave the discussion of it to another occasion.