Gravity and Gauge Theory*

Steven Weinstein
Northwestern University

July 31, 1998

Abstract

Gauge theories are theories which are invariant under a characteristic group of "gauge" transformations. General relativity is invariant under transformations of the diffeomorphism group. This has prompted many philosophers and physicists to treat general relativity as a gauge theory, and diffeomorphisms as gauge transformations. I argue that this approach is misguided.

The theories of three of the four fundamental interactions of nature are "gauge" theories. A central feature of such theories is their invariance under a group of local transformations, i.e., transformations which may vary from spacetime point to spacetime point. The characteristic group of these "gauge" transformations is called the "gauge group."

The theory of the fourth interaction, gravity, is general relativity. General relativity has its own invariance group, the diffeomorphism group. Insofar as one understands "gauge theory" to mean a theory in which "the physics" (more on the ambiguity of this term later) is invariant under a certain group of transformations, one might be tempted to construe general relativity as a gauge theory.[1] Just such a construal figures in recent work of Belot (1996) and Belot and Earman (1997a,b), who follow many (but

*Thanks to Arthur Fine, Chris Isham, and Bob Wald for helpful comments.

[1] One can also regard gauge-invariance (or for that matter diffeomorphism-invariance)

not all) physicists in treating the diffeomorphism group as a gauge group, and who draw implications for the "hole argument."[2] In this paper, I show that general relativity is not a gauge theory at all, in the specific sense that "gauge theory" has in elementary particle physics. This issue is of crucial importance to attempts to quantize general relativity, because in quantum theory, the generators of gauge transformations are emphatically *not* treated as observables, while the generators of spatiotemporal (e.g., Lorentz) transformations are in fact the canonical observables. Thus the discussion in this paper sheds light on the origin of some of the deep and longstanding difficulties in quantum gravity, including the "problem of time", a familiar form of which arises from treating the parametrized time-evolution of canonical general relativity as a gauge transformation.[3]

1 What is a gauge theory?

The standard model of particle physics is made up of three gauge theories (or two, depending on how one counts), corresponding to three of the four known ways in which matter and fields interact. The strong, electroweak, and electromagnetic forces are respectively characterized by the local ("internal") symmetry groups $SU(3)$, $SU(2) \times U(1)$ and $U(1)$.

The power of the principle of gauge invariance is that it effectively determines the structure of the various interactions, so that guessing a gauge group is tantamount to guessing a theory. The theories arrived at through this procedure have led to surprisingly accurate predictions, so not only does the gauge concept constrain the structure of theories, it actually seems to

as a property of the equations of motion. I.e., a theory is gauge- (diffeomorphism-) invariant if a solution of the equations of motion, when acted on by a gauge transformation (diffeomorphism) yields another solution. But this rather formal criterion is misleading, as we shall see in the penultimate section.

[2] The discussion of the hole argument in the philosophy literature began with Earman & Norton (1987). Stachel (1994) is a useful starting point for the uninitiated.

[3] See Isham (1993) or Kuchař (1992) for a review of the problem of time.

lead to *correct* theories! For our investigation into gravity and gauge, we will want to get a handle on the role of the gauge group. Let us begin by examining the way in which the postulation of an invariance under local $U(1)$ transformations leads to Maxwell's theory. We will shortly place these ideas in the more general mathematical context of fiber bundles.

Suppose we have a single, free non-relativistic particle described by a wave-function $\psi(\vec{x})$. Multiplying this wave function by a complex number of unit modulus (a member of the group $U(1)$, e.g., a phase factor of the form $e^{i\theta}$) yields a wave-function that is physically equivalent to the original. The probability distributions for position and momentum (and thus all other observables) are the same, and the time-evolution of the probability distributions is the same. So we say that ordinary quantum mechanics is invariant under a global $U(1)$ transformation, "global" meaning "everywhere in space at some time."

Now suppose we want to make a local, "position" dependent change in the wave-function, i.e., we want to allow the phase at one point in space to differ from that at another point. This implies that we consider transformations of the form $\psi'(\vec{x}) = e^{i\theta(\vec{x})}\psi(\vec{x})$. Then the transformed states will *not* be equivalent, for although they will have the same probability distribution for position, they will have a different probability distribution for momentum (because the phase of a wave-function in configuration space effectively encodes the momentum of the particle), and different dynamical evolution, to wit:

$$
\begin{aligned}
i\hbar\frac{d}{dt}\psi'(\vec{x}) &= \widehat{H}\psi'(\vec{x}) && \text{(1a)}\\
&= \frac{\hat{p}^2}{2m}\psi'(\vec{x}) && \text{(1b)}\\
&= \frac{-\hbar^2}{2m}\vec{\nabla}^2(e^{i\theta(\vec{x})}\psi(\vec{x})) && \text{(1c)}\\
&= \frac{-\hbar^2}{2m}\left[(\vec{\nabla}^2 e^{i\theta(\vec{x})})\psi(\vec{x})) + e^{i\theta(\vec{x})}(\vec{\nabla}^2\psi(\vec{x}))\right] , && \text{(1d)}
\end{aligned}
$$

whereas

$$i\hbar \frac{d}{dt}\psi(\vec{x}) = \widehat{H}\psi(\vec{x}) \tag{2}$$

$$= \frac{-\hbar^2}{2m}\vec{\nabla}^2\psi(\vec{x}) \ . \tag{3}$$

The first term on the right-hand side of (1d) vanishes if and only if $\vec{\nabla}\theta(\vec{x}) = 0$, i.e., only in the case of a *global* phase transformation. In such a case, $d\psi/dt = d\psi'/dt$, modulo a global (hence physically meaningless) phase factor. Otherwise, for local phase transformations ($\vec{\nabla}\theta(\vec{x}) \neq 0$), we find that $d\psi/dt \neq d\psi'/dt$.

Now the way in which the demand for gauge invariance dictates the dynamics of the theory is that we require that the Hamiltonian $\widehat{H}$ be such that $\widehat{H}\psi = \widehat{H}\psi'$. This is accomplished by a Hamiltonian of the form $\widehat{H} = \frac{1}{2m}(-i\hbar\vec{\nabla} + ie\vec{A})^2 - e\phi$, where $\vec{A}$ and ϕ turn out to be, respectively, the vector and scalar potentials of electromagnetism, and where e is the electromagnetic coupling constant, the value of which must be determined by experiment.[4]

Two caveats before we move on. The demand for local gauge invariance is rather undermotivated in the case of a single (or even multiple) particles, and in fact the idea that we perform local gauge transformations by operating on a single particle wavefunction tends to sidestep the awkward fact that the wavefunction is a function on configuration space, which only in the case of an unconstrained, single particle happens to coincide with (three-dimensional) physical space. On the other hand, if one begins with the Klein-Gordon equation or Dirac equation to describe matter, one begins with the concept of a matter *field*, which has an infinite number of degrees of freedom (two per point in space). Since the values of the field at spacelike separated points may be specified independently of one another, it makes sense to conjecture that the phase at each point be independent.[5] Thus a

[4]The electromagnetic coupling constant is one of the "free parameters" in the standard model of particle physics.

[5]It is notable, however, that in a covariant formulation, the phase changes at causally

rather arbitrary requirement in the case of a single particle is much better motivated when one considers a field.

The second caveat is that the demand for gauge invariance does *not* give us the equations for the free electromagnetic field—it only gives us the interaction of the field with the charge (hence the Schrödinger equation for the charge). The requirements of relativistic invariance and renormalizability are additional requirements that do, however, uniquely constrain the field equations, at least in the case of electrodynamics (Schwinger (1953)).

To place gauge theory in a more general perspective, it is helpful to consider the fiber bundle formulation. A fiber bundle is a structure (E, π, M), where E is the "total space", consisting of all the fibers (they form a manifold) and M is the "base space" (also a manifold). The projection map $\pi : E \rightarrow M$ associates each fiber with a point in the base space. The idea in gauge theory is to consider group bundles, where each fiber is a copy of the relevant internal symmetry group, and where the base space corresponds to spacetime.[6] Thus the inverse π^{-1} of the projection map associates a copy of the internal symmetry group (a fiber) with each point in spacetime. For the $U(1)$ gauge theory, the fibers are the group $U(1)$, and each point on a fiber corresponds to a different element $e^{i\theta}$ of $U(1)$. A choice of gauge (i.e. a choice of phase at each point) then corresponds to a cross-section s of the bundle (figure 1).

In this fiber bundle picture, the vector potential A_α (in its 4-dimensional version, where the scalar potential $\phi = A_0$ plays the role of the fourth, "time" component) enters as the connection on the bundle.[7] A connection

connected (timelike or null) points are typically regarded as independent, despite the fact that these fields are in causal contact.

[6]The gauge group ${}^{M}U(1)$ is the set of functions from M into G, meaning that an element of the group is an object of the form $U(1) \times M$, which associates an element of $U(1)$ to every point in spacetime. It is the automorphism group of the principal bundle. (See Göckeler & Schücker (1987).)

[7]More accurately, it is a "correction" to the standard "flat" connection. The standard connection takes one from a given phase $e^{i\phi}$ on one fiber to the same phase $e^{i\phi}$ on the

Figure 1: $U(1)$ fiber bundle

is used to compare vectors at one point with vectors at another, so that one can meaningfully talk about the rate of change of vectors from one point to another. The vectors in the $U(1)$ case are just the local phases. If the connection is such that the phase undergoes a change when moving around a closed loop, this means that the "curvature" $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha$ of the connection A_α is non-zero in the vicinity of the loop, i.e., that there is an electromagnetic field there.

Gauge-invariance is realized because, although the connection changes under gauge transformations, the physical quantities, which are represented by the curvature $F_{\alpha\beta}$, do *not*.[8] Supposing the phase changes by angle θ at each point (i.e., an addition of $e^{i\theta}$), the connection becomes $A'_\alpha = A_\alpha - \partial_\alpha \theta$. The curvature of this new connection is the same as that of the old (because

neighboring fiber. Baez & Muniain (1994) is a useful reference.

[8]Note, however, that the situation is not quite as simple in (non-Abelian) Yang-Mills theory. There, the curvature of the connection is *not* invariant under gauge transformations.

$\partial_\alpha \partial_\beta \phi - \partial_\beta \partial_\alpha \phi \equiv 0$), and so this fiber-bundle formulation nicely embodies the idea of a theory in which the physics (here encoded in the curvature $F_{\alpha\beta}$) is invariant under local $U(1)$ gauge transformations. Notice that what we are saying here is that two vector potentials A_α and A'_α give rise to the *same* tensor $F_{\alpha\beta}$, where "same" means "takes the same values at every point of the manifold M."[9] $F_{\alpha\beta}$ is known as the "Maxwell tensor."

In the Cartesian coordinate system adapted to an observer whose world-line has tangent vector $u^\alpha = (1, 0, 0, 0)$, the Maxwell tensor is

$$F_{\alpha\beta} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} . \tag{4}$$

These quantities are the values of the electric and magnetic fields in the observer's reference frame. As we have seen, they are gauge-invariant. They correspond to the local observables $E_i(\vec{x}, t)$ and $B_i(\vec{x}, t)$ of classical Maxwell theory, and as one would expect they are represented by self-adjoint operators $\hat{E}_i(\vec{x}, t)$ and $\hat{B}_i(\vec{x}, t)$ in the canonical quantum theory.[10]

[9] The physics of this theory is not *entirely* contained in the curvature of the connection. The Aharonov-Bohm effect is an observable effect in regions where there is no electromagnetic field ($F_{\alpha\beta} = 0$), though it does rely on there being a field *some*where. Essentially, one can have a situation in which particles that start at the same place (with same initial phase), traverse different trajectories through regions in which the connection is "flat", but where the particles are nonetheless out of phase when they meet. This is entirely analogous to the failure of parallel transport on the surface of a cone. The tip of the cone corresponds to the region of non-zero electromagnetic field. See Bernstein & Phillips (1981) for an excellent, non-technical exposition.

[10] Technically, they are operator-valued distributions, which must be smeared by appropriate "test functions" to produce true self-adjoint operators.

2 General relativity

General relativity is a theory in which spacetime is represented by a four-dimensional differentiable manifold (a collection of smoothly connected points) equipped with a Lorentz metric giving the spatiotemporal distances between points. "Models" thus consist of a manifold, a metric, and optionally one or more matter fields; what makes such a collection a model of general relativity is that it satisfies Einstein's equation. The particular placement of the metric and other fields on the manifold is arbitrary, reflecting the fact that the physical content of the theory consists in its descriptions of correlations between the various fields. Different placements are related by diffeomorphisms, and the indifference of the physical predictions to diffeomorphisms leads the theory to be characterized as "diffeomorphism invariant."[11]

Perhaps the first thing to note about general relativity is that the diffeomorphism group is not the characteristic group of a group bundle at all. Recall that in a gauge theory, a copy of the gauge group sits over each point of the manifold, and gauge transformations induce changes in the connection (the gauge field) at each point, changes that nonetheless leave the physics at the point unchanged. Diffeomorphisms are a completely different sort of beast—there is no copy of the diffeomorphism group sitting over each point of the manifold, nor is the diffeomorphism group the product $G \times M$ of some other internal group G with the manifold M.

Rather than inducing changes in the field at a point, diffeomorphisms map points x to other points x', and induce changes in the fields (metric tensor, Riemann tensor, etc.) by mapping the fields from one point to another. As a consequence, the value of a field at a given point is physically meaningless in a diffeomorphism-invariant theory.

It is clear, then, that the diffeomorphism group is unlike an ordinary gauge group in that it is not a "local" (in the sense of "at the same manifold

[11] It is in fact not entirely clear that two models related by a diffeomorphism that is not homotopic to the identity are indeed physically equivalent.

point") transformation. Nonetheless, many physicists and philosophers take the view that there is a relevant analogy between the diffeomorphism group and gauge groups. Perhaps the clearest expression of this view comes from Wald:

> If a theory describes nature in terms of a spacetime manifold, M, and tensor fields, $T^{(i)}$, defined on the manifold, then if $\phi : M \to N$ is a diffeomorphism, the solutions $(M, T^{(i)})$ and $(N, \phi^* T^{(i)})$ have physically identical properties. Any physically meaningful statement about $(M, T^{(i)})$ will hold with equal validity for $(N, \phi^* T^{(i)})$. On the other hand, if $(N, T'^{(i)})$ is not related to $(M, T^{(i)})$ by a diffeomorphism and if the tensor fields represent measurable quantities, then $(N, T'^{(i)})$ will be physically distinguishable from $(M, T^{(i)})$. Thus, the diffeomorphisms comprise the gauge freedom of any theory formulated in terms of tensor fields on a spacetime manifold. In particular, diffeomorphisms comprise the gauge freedom of general relativity. (Wald (1984), p. 438)

The idea is that models that differ by a diffeomorphism encode the same physical predictions, just as do models of a gauge theory that differ by a gauge transformation. The essential difference, as we shall see in a moment, is in what sorts of things count as "physical predictions."

First, though, note that many, many theories, including Maxwell theory and other gauge theories, can be (and in fact routinely are) formulated in terms of tensor fields on a spacetime manifold. So, on the face of it, Maxwell theory is just as diffeomorphism-invariant as general relativity. However, we do not tend to treat it as such. That is, we consider quantities such as the components of the Maxwell tensor at a spacetime point x to be observables (physical predictions of the theory) even though such quantities are not diffeomorphism-invariant. The reason for this is that spacetime points x are understood to acquire physical meaning through some implicit background structure. This is of course possible because of the intimately related facts that (a) there are physical objects that do not couple to the electromagnetic field (namely, those objects without electric charge), and (b) the electromagnetic field does not describe the structure of spacetime itself. Thus we can use non-dynamical (here, uncharged) objects to define a spatiotemporal reference frame. Because these objects are not dynamical, we do not include them in the theory.

General relativity is different. All objects are sources of gravity, and therefore all objects are explicit in a theory of gravity. Since the reference system is explicit, the "physically meaningful statements" of general relativity are fundamentally *relational* in nature. For example, if we have a general relativistic model $(M, g_{\alpha\beta}, F_{\alpha\beta})$, where $g_{\alpha\beta}$ is the metric tensor and $F_{\alpha\beta}$ is the Maxwell tensor, then a typical prediction of the model will be of the form "the curvature of spacetime is so-and-so where the electromagnetic field has such-and-such value". Because the fields all transform together under diffeomorphisms, these sorts of predictions are, indeed, invariant under diffeomorphisms.

In short, the physical predictions of a gauge theory, classical or otherwise, are predictions about the values of fields at spacetime points specified with respect to a fixed background spacetime. These are physically meaningful because it is assumed that one has some structure external to the model which identifies these points. In a diffeomorphism invariant theory such as general relativity, however, there *is* no background structure, and so the "physically meaningful statements" take on a purely relational character.[12] The upshot is that when one talks about the physical predictions being invariant under a gauge transformation, one is talking about a very different sort of thing than when one talks about the physical predictions being invariant under a diffeomorphism.

In closing, it is worth mentioning another sense in which general relativ-

[12]The exceptions to this statement are certain "global" quantities such as the "ADM mass" which are defined at spatial infinity in asymptotically flat spacetimes.

ity can be understood as a gauge theory. Various authors have attempted to derive general relativity from a gauge-like principle, involving invariance of physics under transformations of the locally acting (i.e. in the tangent space at each point) Lorentz or Poincaré group. (See for instance Utiyama (1956), Kibble (1961) and more recently Wilczek (1998).) Though some of these theories are interesting in their own right, in none of them is the diffeomorphism group a gauge group.

3 Conclusion

The diffeomorphism group is simply not a gauge group in the specific sense this term has in particle physics. Formally, the distinction is clear enough—the diffeomorphism group is not the automorphism group of a principle fiber bundle. Physically, though there is a sense in which "the physics" of a gauge theory and the physics of a diffeomorphism-invariant theory is respectively invariant under gauge transformations and diffeomorphisms, what is *meant* by "the physics" is quite different.

The distinction between the two sorts of invariance is absolutely vital in the context of quantum theory. We understand how to quantize gauge theories. A primary aspect of this is to represent some subset of the classical gauge-invariant quantities (classical observables) as self-adjoint operators (quantum observables) obeying relevant commutation relations. However, it is not at all clear what it means to quantize a diffeomorphism-invariant theory. One can attempt to turn the classical diffeomorphism-invariant quantities into observables—this is effectively what one is doing when one formally treats the diffeomorphism group as a gauge group. However, in addition to leading to the problem of time in canonical gravity (Isham (1993), Kuchař (1992), Weinstein (1998*a*)) and possibly a breakdown of the superposition principle (Weinstein (1998*b*)), this approach leaves one with virtually no known observables *whatsoever* for vacuum gravity in the compact case (Torre (1993), Torre (1994)). Indeed, it seems premature to apply such techniques to general relativity—one would like to see a diffeomorphism-invariant version of a simpler theory (e.g. Maxwell theory) first.[13] That this has only rarely been attempted is perhaps explained by the inclination of many physicists working in the area to think of the diffeomorphism group as a gauge group, in particular as the special gauge group characteristic of general relativity. But this is clearly misguided.

References

Baez, J. & Muniain, J. (1994), *Gauge Fields, Knots and Gravity*, World Scientific, Singapore.

Belot, G. (1996), Whatever Is Never and Nowhere Is Not: Space, Time, and Ontology in Classical and Quantum Gravity, PhD dissertation, University of Pittsburgh.

Belot, G. & Earman, J. (1997*a*), 'From physics to metaphysics'. To appear in the Festschrift for Michael Redhead, ed. H. Brown and R. Clifton.

Belot, G. & Earman, J. (1997*b*), 'Pre-Socratic quantum gravity'. To appear in *Philosophy at the Planck Length*, ed. C. Callender and N. Huggett.

Bernstein, H. & Phillips, A. (1981), 'Fiber bundles and quantum theory', *Scientific American* **245**, 94–109.

Earman, J. & Norton, J. (1987), 'What price spacetime substantivalism?', *British Journal for the Philosophy of Science* **38**, 515–525.

Göckeler, M. & Schücker, T. (1987), *Differential Geometry, Gauge Theories, and Gravity*, Cambridge University Press, Cambridge.

[13] See Rovelli (1995) for an attempt at constructing a diffeomorphism-invariant quantum field theory.

Isham, C. (1993), Canonical quantum gravity and the problem of time, *in* 'Integrable Systems, Quantum Groups, and Quantum Field Theories', Kluwer Academic Publishers, London, pp. 157–288.

Kibble, T. (1961), 'Lorentz invariance and the gravitational field', *Journal of Mathematical Physics* **2**, 212–221.

Kuchař, K. (1992), Time and interpretations of quantum gravity, *in* 'Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics', World Scientific, Singapore, pp. 211–314.

Rovelli, C. (1995), 'Outline of a generally covariant quantum field theory and a quantum theory of gravity', *Journal of Mathematical Physics* **36**, 6529–6547. gr-qc/9503067.

Schwinger, J. (1953), 'The theory of quantized fields II', *Physical Review* **91**, 713–728.

Stachel, J. (1994), The meaning of general covariance: the hole story, *in* J. Earman, A. Janis, G. Massey & N. Rescher, eds, 'Philosophical Problems of the Internal and External Worlds', Pittsburgh-Konstanz Series in the Philosophy and History of Science, University of Pittsburgh, Pittsburgh, pp. 129–160.

Torre, C. (1993), 'Gravitational observables and local symmetries', *Physical Review* **D48**, 2373–2376.

Torre, C. (1994), 'The problems of time and observables: some recent mathematical results'. gr-qc/9404029.

Utiyama, R. (1956), 'Invariant theoretical interpretation of interaction', *Physical Review* **101**, 1597–1607.

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

Weinstein, S. (1998*a*), Conceptual and Foundational Issues in the Quantization of Gravity, PhD dissertation, Northwestern University.

Weinstein, S. (1998*b*), Time, gauge, and the superposition principle in quantum gravity, *in* 'Proceedings of the Eighth Marcel Grossman Meeting', World Scientific, Singapore. gr-qc/9711056.

Wilczek, F. (1998), 'Riemann-Einstein structure from volume and gauge symmetry'. hep-th/9801184.