

# On the Reality of Gauge Potentials\*

Richard Healey

Philosophy Department, University of Arizona

---

\*

Send requests for reprints to the author, Department of Philosophy, University of Arizona, Tucson, AZ 85721-0027; email, rhealey@U.Arizona.edu.

I thank Frank Arntzenius, Jeff Barrett, Jenann Ismael, Steven Leeds, David Malament, Tim Maudlin, Paul Teller; several referees for *Philosophy of Science*; and especially Harvey Brown, who introduced me to the work of Jeeva Anandan.

## Abstract

Classically, a gauge potential was merely a convenient device for generating a corresponding gauge field. Quantum-mechanically, a gauge potential lays claim to independent status as a further feature of the physical situation. But whether this is a local or a global feature is not made any clearer by the variety of mathematical structures used to represent it. I argue that in the theory of electromagnetism (or a non-Abelian generalization) that describes quantum particles subject to a classical interaction, the gauge potential is best understood as a feature of the physical situation whose global character is most naturally represented by the holonomies of closed curves in space-time.

**1. Introduction.** Three of the four fundamental interactions acknowledged today are currently best described by quantized gauge theories. A standard formulation of such a gauge theory incorporates mathematical objects representing gauge potentials over and above the related objects that represent the gauge fields these generate. But how seriously should we take this aspect of the mathematics? Should the mathematical object(s) typically called the gauge potential be taken to represent anything physically real if the theory is true, or are they there just to grease the calculational wheels? If this mathematical gauge potential is understood to represent something physically real, then is that representation relation one-one or many-one, and what, if anything, can be said without using this representation about the structure so represented? Current gauge theories of the electromagnetic, weak and strong interactions are quantized, so these questions cannot be fully answered independently of an interpretation of quantum field theory. But since that project raises apparently unrelated difficulties it seems a wise strategy to treat these only after gaining some understanding of gauge potentials prior to quantization. The present paper is therefore restricted to the context of classical gauge theories, and specifically to an examination of the status of gauge potentials in classical electromagnetism and its non-Abelian generalizations (Yang-Mills theories), where these are taken to influence the behavior of quantum-mechanically described particles. The relation

between mathematical and physical structures is problematic enough even in this context. I shall argue that, while interference phenomena show that the locally defined mathematical objects typically called gauge potentials do indeed represent something physically real in each of these theories, the physical structure they represent itself has a distinctly global character. The case against local physical gauge potentials is not so different from the case against absolute space. Both cases deploy epistemological principles of wide applicability in theoretical science.

**2. Are Gauge Potentials Real?**<sup>1</sup> Potentials have long played a role in electromagnetism, the first gauge theory. After the demise of attempts to find mechanical models of Maxwell's equations, it became widely accepted that the so-called scalar electric potential and magnetic vector potential were just calculational aids in classical electromagnetism, with no physical significance independent of the electric and magnetic fields they helped one to calculate. The advent of relativity made it natural to combine them into the electromagnetic (EM) 4-vector potential  $A^\mu$  (or  $A_\mu$ ). Mathematically, this is a vector field--a smooth map from a space-time manifold into its tangent (or cotangent) spaces. Its role was thought to be exhausted by its use to derive the EM field, now represented by the tensor  $F_{\mu\nu}$  (whose nonzero components in a particular Lorentz frame may be identified with those of the electric and magnetic fields in that frame), via the equation  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ . Now this relation is invariant under the *gauge transformation*

$$A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu \Lambda(x) \quad (1)$$

for a suitably differentiable, but otherwise arbitrary, function  $\Lambda(x)$ . It follows that there is an infinite class of 4-vector potentials  $A_\mu$  that generate the same field  $F_{\mu\nu}$ , none of which were taken to represent any additional physical structure.

Things changed with the advent of wave mechanics. Electromagnetism entered into mechanics now by the introduction of  $A_\mu$  directly into the Hamiltonian operator appearing in the Schrödinger equation, rather than indirectly via the presence of  $F_{\mu\nu}$  in the classical Lorentz force law. In their seminal 1959 paper, Aharonov and Bohm pointed out the implication that the effects of electromagnetism on a beam of charged particles that passes through a region of space are not always wholly determined by the electromagnetic field there while it passes. Even for a fixed electromagnetic field in this space-time region, the location of fringes in an interference pattern formed by the beam may depend on the electromagnetic field in space-time regions from which the beam is excluded. Rejecting action at a distance, Aharonov and Bohm themselves took this phenomenon to demonstrate the physical reality of the electromagnetic potential in the region accessible to the beam, and indeed more generally. They were followed here by Feynman 1965 in his influential lectures. Different interference patterns corresponded to different 4-vector fields  $A_\mu$  (but the same electromagnetic field  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ ) in the space-time region accessible to the beam that produced them. But it was well known that not only the electromagnetic field but also the observable interference pattern remain unchanged under the gauge transformation (1).

This is puzzling. Interference phenomena provide evidence that there is more to electromagnetism than is represented by the field  $F_{\mu\nu}$ , but there is nothing in these phenomena or in the theory that accounts for them to single out any particular  $A_\mu$  as faithfully representing this additional physical structure. In this paper I shall use the term ‘electromagnetic potential’ to refer to such a physical structure (if it exists), not to any particular mathematical representation of it. Similar physical structures apparently figure in non-Abelian generalizations of EM theory. I shall refer to all of these structures as gauge potentials. The general problem then is this. Phenomena like the Aharonov-Bohm effect are naturally taken to provide evidence that gauge potentials are real physical structures, once one rules out gauge fields that act at a distance. But what kind of structures are they, and how is their character best represented mathematically? I shall evaluate several answers to these questions before proposing another.

**2. Are Gauge Potentials Real Vector Fields?** A simple mathematical representation of a gauge potential is that of the EM potential by a 4-vector field  $A_\mu$ . This mode of representation generalizes naturally to other gauge theories. For example, the Yang-Mills potential for an SO(3) gauge theory may be represented by a 4-vector field  $\mathbf{W}_\mu$ , written boldface to indicate that it is also a vector in the "internal" SO(3) space--the space of 3D Euclidean rotations. This prompts the following interpretative suggestion. Physically, gauge potentials are structured sets of properties (4-vector quantities) of space-time points: the properties of a region are faithfully

represented by a unique 4-vector field on that region. On this suggestion, the EM gauge potential in a region is faithfully represented by a particular 4-vector field  $A_\mu$  defined on that region. The representation is faithful in that to each possible set of EM gauge properties at each space-time point  $x^\mu$  there corresponds one and only one value  $A_\mu(x^\mu)$  of  $A_\mu$ : a non-trivial gauge transformation (1) applied to  $A_\mu$  yields a 4-vector field  $A_\mu'$  that misrepresents the gauge potential in the region. (In a non-Abelian theory the set of locally-defined gauge properties would correspond to a magnitude with direction in the corresponding "internal" space as well as in external 4D space-time.)

In one sense, such locally-defined structured sets of properties are quite familiar. The vector field for the 4-momentum density of a continuous cloud of dust would represent a similarly structured set of locally-defined properties. But while the mode of representation is quite familiar, the particular sets of properties supposedly represented here are not. What is being suggested is that, in addition to its properties corresponding to the local electric and magnetic fields (or their non-Abelian analogues), each space-time point has a further set of gauge properties. While these have no influence on the trajectories of classical particles, they do act locally on the phase of the wave-functions of charged particles, and may thereby affect interference phenomena in which these are involved. If this suggestion is correct, then we can use the unique representation by a 4-vector field to refer to these gauge properties of each space-time point, and at least to that extent say what those properties are.

But the suggestion faces severe difficulties.<sup>2</sup> Consider a restricted region of space-time

in which are present a particular gauge field and gauge potential. Call this a particular *gauge situation*. Suppose that one such gauge situation actually occurs during the performance of an experiment to verify the existence of the Aharonov-Bohm effect: in that situation the EM field is everywhere zero, but the EM potential is represented by a 4-vector field that cannot be transformed to zero everywhere by a gauge transformation. On the present suggestion, exactly one 4-vector field faithfully represents the gauge properties within the region in question: this is the 4-vector field in the ONE TRUE GAUGE. Nontrivial gauge transformations yield other 4-vector fields that represent gauge situations other than the actual one. But those alternative gauge situations are empirically indistinguishable from the actual situation: they feature indistinguishable interference effects, as well as indistinguishable trajectories for classical charged particles.

Now consider some particular 4-vector field  $A_\mu$  that is compatible both with the EM field (as revealed by the trajectories of classical particles) and with the observed interference effects in the actual situation. Does  $A_\mu$  represent the EM potential in the actual gauge situation, or does it rather represent the EM potential in some other gauge situation that is merely experimentally indistinguishable from the actual situation? This question is pressing, but no experiment is any help in answering it. On this suggestion, gauge theories are afflicted by radical underdetermination. In this and every other gauge situation we cannot acquire any evidence relevant to determining which of an infinite class of gauge-equivalent 4-vector fields faithfully represents the real gauge potential in that situation.



That this is so should not be surprising. For it is the universal practice among physicists to treat the choice of gauge as purely conventional. In this respect, gauges are like coordinate systems. The suggestion that there is ONE TRUE GAUGE in which the 4-vector potential faithfully represents the actual gauge situation is like the suggestion that there is ONE TRUE COORDINATE SYSTEM for faithfully describing what happens during any particular experiment.

There is another objection to the suggestion that the gauge potential in a particular gauge situation is faithfully represented by a single, privileged, 4-vector field. Suppose one somehow knew (by divine revelation?) the values of this 4-vector field over part of the space-time region in question--say, for the initial half of the region, relative to some Lorentz frame. This would still leave open an infinite number of possibilities for the development of this field in the rest of the region. For the future development of the 4-vector field in a region is not determined by its values on a time-slice in that region, or even on its values to the past of such a slice, no matter how far the region extends in space and time. This is a simple consequence of the fact that, among the gauge transformations (1), there is a continuous infinity of transformations that are the identity within a given region, but diverge smoothly from the identity outside that region.

**3. Fiber Bundle Formulation of Gauge Theories.** A plausible diagnosis of the origin of these problems facing the idea that a gauge potential is faithfully represented by a particular 4-

vector field is that this idea violates gauge invariance. The gauge transformation (1) takes the one 4-vector field that (supposedly) faithfully represents the physical gauge potential into other 4-vector fields that misrepresent it, even though the description of the observable phenomena is invariant under (1). Now since the pioneering work of Wu and Yang 1975 it has become common practice for mathematical physicists to formulate gauge theories within the richer mathematical framework of fiber bundles. And this may seem to promise a natural way of ensuring gauge-invariance by offering an alternative mathematical representation of a gauge potential by the connection on a principal fiber bundle.

In the next section I question whether this alternative representation warrants an interpretation of a gauge potential as a locally defined structure. The present section lays out the framework on which such an interpretation rests. Readers impatient with the technical details are encouraged to skip it, especially since the main positive thesis of the paper can be stated and defended independently of the fiber bundle formulation. But the charms of this formulation of gauge theories have captivated not only mathematical physicists but also some philosophers, who are unlikely to be convinced of that thesis if these details are ignored.

In the fiber bundle formulation, the EM potential is represented not by any particular vector field on a space-time manifold, but by a connection on a principal fiber bundle which has the space-time manifold as its base space. A choice of gauge corresponds to a choice of section for the bundle. To each local section of the bundle there corresponds a unique vector field  $A_\mu$  on an open set of the base space that represents the potential relative to that section. Different

bundle sections yield different vector fields related on their common domains by the gauge transformation (1). None of these provides a gauge-invariant representation of the physical gauge potential. The EM field is represented by the curvature of the connection. The state of quantum particles of charge  $q$  subjected to electromagnetic effects is represented on another associated vector bundle. The structure group that defines the fibers of both bundles is  $U(1)$ --the set of rotations in the complex plane, parametrized by the angle of rotation. In any principal bundle, the elements of the fiber are just the members of the structure group itself, whereas the fiber of the associated vector bundle consists of a vector representation of that group. In the case of electromagnetism, a vector in a fiber of the associated vector bundle is just a complex number. This is the value of the particles' position-representation wave-function at the space-time point lying "below" that fiber. A change of phase of the wave-function at that point corresponds to a rotation in the fiber "above" that point. There are in fact many different position representations, related to one another by unitary Hilbert space transformations.<sup>3</sup> A local phase transformation--a rotation in the fiber "above" each point that varies smoothly from point to point--corresponds to a change from one position representation to another. Wave-functions related by a local phase transformation can therefore be considered to be just different representations of one and the same abstract Hilbert space state vector. A wave-function in a particular position representation is a section of the vector bundle--a smoothly varying choice of one element from the fiber "above" each point of space-time. It corresponds uniquely to a section of the principal fiber bundle representing electromagnetism.

The connection on the principal fiber bundle representing electromagnetism is a geometric object: specifically, it is given by a Lie-algebra-valued one-form field on the bundle. It is defined independently of any choice of coordinate charts or section for the bundle. The pull-back corresponding to each local section on this bundle uniquely defines a one-form (or co-vector) field on a corresponding open set of the (space-time) base manifold  $M$ . The usual quantity  $A_\mu$  is just a coordinate representation of this one-form field, but for a trivial constant factor. The usual quantity  $F_{\mu\nu}$  is similarly related to a coordinate representation of the two-form field on  $M$  given by the pull-back of the bundle curvature. In the absence of magnetic monopoles there are global sections for which the set on which the one-form field is defined encompasses the entire manifold. In that case a coordinate chart may be chosen for the base manifold so that this field is defined everywhere in space-time. A change in space-time coordinates  $x \rightarrow x'$  will merely change the coordinate representation  $A_\mu(x) \rightarrow A'_\mu(x')$  without changing the one-form field on the space-time manifold this represents. A change in choice of section for the fiber bundle, on the other hand, will change this one-form field itself. Such a change corresponds to what physicists call a local gauge transformation of the electromagnetic potential. It corresponds uniquely to a change of section on the associated vector bundle on which is represented the state of the particle. This change of section induces a shift from one position representation of the quantum state to another--what physicists call a local gauge transformation of the quantum state of the particle. A local gauge transformation can therefore be described in a particular space-time coordinate chart by the simultaneous transformations

$$\Psi(x) \rightarrow \exp(i\Lambda(x)) \Psi(x) : \quad A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu \Lambda(x) \quad (2)$$

The fiber bundle formulation is more general than the older formulation, since it allows for the possibility of a gauge potential that cannot be represented as a single one-form field over the entire space-time manifold. As Wu and Yang(1975) show, this would be the case in the presence of a magnetic monopole. Even in such a case, the electromagnetic potential is uniquely representable as the connection of a principal fiber bundle. But because this bundle is nontrivial, it has no global sections. While each local section yields a one-form field representing the electromagnetic potential on a corresponding open set of the space-time manifold, there are different one-form fields on different open sets (related on their overlaps by so-called transition functions). In the absence of convincing evidence that there are any magnetic monopoles, the employment of the fiber bundle formulation of electromagnetism is regarded by many physicists as a mathematically elegant option to the older treatment with no deeper physical significance. Multiplying the entire wave-function by a constant phase factor  $\exp i\phi$  yields a wave-function that is standardly taken to represent, in the same position-representation, a different vector from the same Hilbert space ray that specifies the system's quantum state. But what the fiber bundle formulation shows is that the result of multiplying the wave-function even by a position-dependent phase factor  $\exp i\Lambda(x)$  may alternatively be regarded as just a different position-representation of one and the same abstract Hilbert space state vector, provided that one simultaneously performs a corresponding transformation on the one-form field on the space-time manifold that represents the electromagnetic vector potential,

in accordance with equation (2).

This local gauge transformation does not correspond to a mere change of space-time coordinates, but to a change in the geometric object on the space-time manifold. But the change from the original to the new one-form field appears importantly analogous to a change of coordinates, in so far as it produces an alternative representation of the very same physical situation. Choosing one of these allowable representations--i.e. a particular position representation for the quantum state-vector, and a corresponding gauge for the electromagnetic potential--comes down to choosing a related pair of sections of the fiber bundles on which the potential and quantum state are defined. The fiber bundles themselves represent the potential and quantum state in a more intrinsic--since more representation-independent--way.

Electromagnetism influences the quantum state through a change in the connection that represents it on the principal fiber bundle. For a fixed section of the principal bundle (and the corresponding section on the associated vector bundle), changing the connection changes the wave-function in the same position-representation.

This all generalizes to non-Abelian Yang-Mills theories. The symmetry group of the theory becomes the structure group of the fiber bundle formulation. The gauge potential is represented in the first instance by the connection on a principal fiber bundle over a manifold representing space-time, while the gauge field is represented by its curvature. Relative to a given bundle section, these may also be represented by Lie-algebra-valued one-form and two-form fields (respectively) on (an open subset of) the base space-time manifold. There is a

corresponding section of the associated vector bundle on which is represented the state of particles subject to this gauge interaction. This defines the particles' wave-function as a multi-component (vector or spinor) field on (this subset of) the space-time manifold. In the absence of magnetic monopoles the principal bundle has a global section, for which a gauge-potential one-form and curvature two-form are defined over the whole of the space-time manifold.

**4. Are Gauge Potentials Real Bundle Connections?** The idea that gauge potentials are faithfully represented by 4-vector fields on the space-time manifold  $M$  seemed to founder on the gauge freedom of the field  $A_\mu$ . Within the fiber bundle formulation, such gauge freedom may be associated with the freedom to choose a section for the principal fiber bundle on which the gauge potential and field are represented as connection and curvature respectively. Since the connection and curvature are mathematical objects that are intrinsic to the bundle, and independent of any choice of section, there is a natural temptation to try to avoid any problems associated with the arbitrariness of choice of gauge by taking these objects instead to faithfully represent corresponding physical structures. This temptation may be particularly forceful in the case of a gauge theory with a non-Abelian structure group such as  $SO(3)$ , in which not only the 4-vector field  $W_\mu$  on  $M$  representing the gauge potential but also the associated tensor field  $W_{\mu\nu}$  on  $M$  representing the gauge field strength depends on a choice of gauge.

The suggestion is that adopting a fiber bundle formulation of a gauge theory is motivated not only by mathematical elegance and openness to future developments like the

discovery of magnetic monopoles, but also because it promises a satisfactory understanding of gauge potentials as (in some sense) locally defined objects. I shall argue that the promise cannot be fulfilled and so the suggestion is fruitless. While a gauge potential may of course be represented by a connection on a principal fiber bundle, this provides no reason to conclude that a gauge potential is in any sense a locally defined structure.

The first problem with the present suggestion is to say just what kind of physical structure could be mirrored by a connection on a principal fiber bundle. The question is, what properties and/or relations of what objects could such a connection faithfully represent? One answer involves fiber bundle substantivalism. On this interpretation, each bundle point represents a different physically real object, and the bundle connection represents a set of intrinsic physical relations among these objects in much the same way that the affine connection on a space-time manifold represents intrinsic geometric relations. It is as if the total space of the bundle represents a higher-dimensional space-time, whose additional dimensions have an affine but not a metric geometry. What we think of as a single point of 4-dimensional space-time represented by point  $p$  of  $M$  is actually a continuum of points represented by the bundle points "above"  $p$ , and the gauge potential is a locally defined object in so far as it characterizes intrinsic relations, in the first instance between *neighboring* bundle points.

This is not a good answer to the question. It yields a luxuriant ontological growth that cries out for treatment by Ockham's razor. A particular supposed object represented by a bundle point  $u$  "above"  $p$  is assigned no properties that could be used to describe or to



distinguish it from another point represented by  $u'$  "above"  $p$ . Unlike points of space-time, points represented by  $u, u'$  cannot be discriminated by ostension, however idealized. On this interpretation, the objects represented by bundle points are mere bearers of the supposed intrinsic relations represented by the bundle connection. Now within the fiber bundle formulation, each interaction described by a gauge potential demands its own principal bundle. This requires many different continua of points lying "above" the same base point  $p$  in different senses of 'above'. The ontological jungle grows thick indeed!

Moreover, the fiber bundle substantivalism involved in this answer is subject to an analogue of the "hole" argument against space-time substantivalism.<sup>4</sup> To set this up, consider once more gauge transformations like (2). When I first introduced these I described them as a mere change in the way a given potential and quantum state are represented, corresponding to a switch from one arbitrarily chosen section and associated position representation to another. Such a change does not reflect any change in the potential or quantum state themselves. If a particular physical situation is specified uniquely by a particular state and potential, then different one-form fields  $A_\mu$  defined on (regions of) space-time merely yield different representations of the electromagnetic aspect of this situation. One might conclude, then, that different one-form fields  $A_\mu$  related by a local gauge transformation do not represent distinct but epistemically indistinguishable states of affairs, but one and the same electromagnetic potential that is *uniquely* represented by a particular connection on a principal fiber bundle.

But there is another way of reading the local gauge transformations (2) according to

which the second of these describes a change in the electromagnetic potential itself—or, more precisely, in the connection that this view supposes to uniquely represent it on the principal fiber bundle. Instead of taking the transformation

$$A_\mu(x) \rightarrow \bar{A}_\mu(x) = A_\mu(x) - \partial_\mu \Lambda(x) \quad (3)$$

to yield an alternative representation of a given fiber bundle connection (corresponding to a new choice of section for the bundle), hold the bundle section fixed and use (3) to define a new fiber bundle connection, whose representation *in the same section* is not  $A_\mu$  but  $\bar{A}_\mu$ . Clearly the old and new fiber bundle connections are distinct geometric objects defined on the total space  $E$ : they are given by distinct Lie-algebra-valued one-form fields on the fiber bundle. One can specify the relation between them intrinsically (without reference to any sections of the bundle) as follows. Let  $f: E \rightarrow E$  be a vertical automorphism of the fiber bundle: a function that maps the total space diffeomorphically onto itself in such a way that it both preserves the fiber “above” each point and respects the action of the structure group on these fiber elements. Then  $f$  induces a map  $f^*$  from the set of connections on  $E$  onto itself by  $f^*(\omega)_{f(e)} = \omega_e$ , where  $\omega_e$  indicates the value of the connection one-form field  $\omega$  at the point  $e$  of  $E$ , and the projection  $\pi: E \rightarrow M$  of the bundle maps both  $e$  and  $f(e)$  onto the same point  $p$  of the space-time manifold  $M$ . The change from the old fiber bundle connection (represented by  $A_\mu$ ) to the new one (represented by  $\bar{A}_\mu$ ) is induced by one such vertical bundle automorphism  $f$ .

Vertical bundle automorphisms corresponding to local gauge transformations generate an infinite class of connections on the bundle, each element of which lays equal claim to

represent an electromagnetic potential. If at most one of these connections is taken to represent the *real* electromagnetic potential, then nothing in the fiber bundle formulation picks out which it is. Moreover, all these connections predict the same observable phenomena--interference effects as well as the trajectories of classical charged particles. This is because in the case of electromagnetism, all observable phenomena involving particles of charge  $q$  are determined by a certain function of the connection, which may be expressed in terms of  $A_\mu$  as the

Dirac phase factor  $\exp[-(iq/\hbar)\int_C A_\mu(x^\mu).dx^\mu]$  where the integral is taken around an arbitrary piecewise differentiable closed curve  $C$  in  $M$ . All connections related by a vertical bundle automorphism are empirically equivalent because the Dirac phase factor for arbitrary  $q$  remains invariant under the transformation (3).

Just as in the original hole argument, the assumption that distinct bundle connections related by a vertical automorphism represent distinct gauge potentials leads to two problems: radical indeterminism and underdetermination. There are vertical automorphisms that diverge continuously from the identity to the future of some time-slice through  $M$ . Hence, even given a complete specification of the connection prior to this time-slice the theory leaves open an infinite number of ways in which the connection (and consequently the gauge potential) can develop subsequently. And the gauge potential is radically underdetermined since no observation is any help in discriminating between supposedly distinct gauge situations represented by distinct connections that yield the same Dirac phase factors for each value of  $q$ .

This analogue to the hole argument need not leave the bundle substantialist speechless.

One reply is *connection essentialism*--the analogue to metric essentialism as a reply to the original hole argument. On this view, what makes a bundle point the point that it is is the way it is related to other bundle points. These relations may be represented by any of the bundle connections that yield the empirically required Dirac phase factors for arbitrary  $q$ . But choosing one such connection to represent the relations will make other empirically equivalent bundle connections represent impossible gauge situations in which bundle points lack some of their essential properties.

Although it does solve the problems of radical indeterminism and underdetermination faced by bundle substantivalism, connection essentialism is a rather unattractive view. The smallest variation on a gauge situation would yield an alternative situation in which the bundle points represent a completely new set of objects. Such modal fragility makes the status of such objects even more tenuous and further undermines the case for their substantiality.

While none of these objections to bundle substantivalism is conclusive, collectively they strongly suggest that the empirical success of a gauge theory like electromagnetism does not warrant the conclusion that gauge potentials are local structures in the sense that they specify intrinsic relations among neighboring objects represented by points of a principal fiber bundle (where each of these objects may itself reasonably be considered "local" in so far as it lies "above" a single space-time point). Ockham's razor will surely rid us of such entities unless it proves impossible to find an alternative interpretation of gauge theory that dispenses with them.

There is an alternative way of implementing the idea that the connection on a principal

fiber bundle directly reflects the structure of the local gauge potential it represents.<sup>5</sup> I shall call it the novel gauge properties view. On this view, bundle points represent intrinsic gauge properties of space-time points, rather than objects "above" those points, while the bundle connection represents relations among these gauge properties. This view cuts down on ontology by substituting additional ideology (in Quine's sense of that term). Moreover, the additional properties are restricted in so far as each space-time point is taken to have just one out of a set of properties with the same structure as the structure group of the gauge theory. In one version, the possible gauge properties of each space-time point are peculiar to that point, and cannot be considered instantiations of universals ranging over all space-time points: the connection then sorts the properties of different points into equivalence classes, but only relative to a particular path linking those points.

In assessing this alternative view, it is important to bear in mind the novelty of the intrinsic gauge properties it postulates. For example, a space-time point is taken to have EM properties over and above those corresponding to the values of the electric and magnetic fields. It should give one pause that no gauge theory mentions such properties, or even contains the vocabulary to do so. Suppose one tried to enrich the language of the gauge theory to permit talk of these novel gauge properties, so that space-time point  $p$  can be said to have novel gauge property  $P$ . The view implies that some other space-time point  $q$  may have one or other of two mutually exclusive gauge properties, say  $Q_1$  and  $Q_2$ . Now consider a curve  $\Gamma$  in  $M$  linking  $p$  to  $q$ . The connection is supposed to represent equivalences among gauge properties of distinct

points, relative to a curve joining them. We may therefore formulate the following question: "Is property  $P$  of point  $p$  equivalent to property  $Q_1$  or property  $Q_2$  of  $q$ , relative to  $\Gamma$ ?" But this question is unanswerable. Nothing in the gauge theory permits one to obtain evidence relevant to deciding which gauge property at one point is equivalent to which gauge property at another point, relative to some path joining these points. This is because the relation of equivalence does no work in the gauge theory--it makes no empirical difference whether two gauge properties are counted as equivalent.

Indeed, the problem is deeper: it is not only epistemological but semantic. If the present view of the structures represented by a gauge theory is correct, nothing in the theory (or elsewhere) makes it possible to refer to particular gauge properties like  $Q_1$  and  $Q_2$  to make contingent claims concerning their possession. So the above question is unanswerable because it has no determinate semantic content. It is no accident that gauge theory gives us no way to talk about the alleged novel gauge properties. Even if there were such properties, one could not set up a coherent referential semantics of a language for ascribing them. But if there are no such novel gauge properties, then there can be no path-dependent equivalence between the properties of distinct space-time points, and no such relation for the connection to represent. I shall outline such an alternative "deflationary" interpretation of gauge theory in the next section. With that in hand, it will be possible to defend its superiority both over bundle substantivalism and over the novel gauge properties view. But there is another objection to the latter view worth raising now.

Consider a non-Abelian gauge theory with no magnetic monopoles. The gauge potential may be represented by a connection on a principal fiber bundle over the space-time manifold  $M$ : the bundle is trivial since it has a global section. Hence one can always represent the gauge potential relative to a global section by a Lie-algebra-valued one-form field defined on the whole of  $M$ . Now one might expect this representation of the gauge potential on  $M$  to be a continuous function of the magnitude of the sources present in a gauge situation. This would generalize the condition that the magnetic vector potential be a continuous function of the current through the solenoid in the magnetic Aharonov-Bohm effect. But the so-called Gribov ambiguity shows that this expectation cannot always be fulfilled.<sup>6</sup> Continuous variations in a non-Abelian gauge situation can take one through a continuous sequence of situations, returning finally to the original gauge situation, in such a way that any local representation of the gauge potential on  $M$  also varies continuously but then *fails* to return to its original form at the end of the sequence.

On the novel gauge properties view, the relations between these properties in a given gauge situation may be faithfully represented by a particular bundle connection. As the situation is varied, the bundle connection will also vary, as will its local representation in a fixed section. But once one has chosen a section, one would expect the local representation of the connection in that section to vary continuously as the gauge situation varies, since this local representation is supposed to mirror the continuously varying relations among the novel gauge properties in these situations. The Gribov ambiguity shows that this cannot always be the case

in a non-Abelian gauge theory. One cannot lay down a convention for representing the supposed relations among novel gauge properties by a bundle connection which satisfies the condition that the local representation of this connection in a fixed section vary continuously with variations in the sources present in a gauge situation. Briefly, but less accurately, gauge potentials cannot be bundle connections because continuous variations in gauge potentials don't always correspond to continuous variations in bundle connections.

This last objection is suggestive but not conclusive. A proponent of the novel gauge properties view can always claim that relations among them may vary continuously with varying gauge situations, while the bundle connection used to represent these relations does not. But the content of this claim is unclear, in so far as our only hold on the supposed relations among novel gauge properties is through their representation by a bundle connection. What is required is some further account of the structure of the relations among novel gauge properties that gives sense to the idea that these vary continuously in continuously varying gauge situations, and explains why it is that the gauge theory's only representation of these relations is unable to mirror this continuous variation. It is difficult to see how the success of a gauge theory could warrant belief in such an account, even if it were forthcoming.

**5. Are Gauge Potentials Holonomies?** Recall that, in the case of electromagnetism, any two bundle connections whose representations in a given section on  $M$  are related by the transformation (3) are empirically equivalent: they yield the same fields as well as the same



Dirac phase factors for arbitrary charge  $q$ . This is an instance of a general result for Yang-Mills gauge theories, though the generalization to the non-Abelian case requires a little care. The key notion is that of the *holonomy* of a connection, which may be explained as follows.

For each piecewise differentiable curve  $\Gamma$  in the base manifold, and each point  $u$  in the fiber  $F_p$  "above" a point  $p$  on  $\Gamma$ , a connection on a principal fiber bundle defines a corresponding curve through  $u$  called the *horizontal lift* of  $\Gamma$ . If  $C_p$  is a closed curve beginning and ending at point  $p$  of the base manifold  $M$ , then its horizontal lift maps each bundle point  $u$  "above"  $p$  into a corresponding point  $v = u_O g$  "above"  $p$ , for some element  $g$  of the structure group  $G$  (see Figure 1). ~~**PLACE FIGURE 1 ABOUT HERE**~~ The set of elements  $g(C_p)$  for all closed curves  $C_p$  forms a subgroup of the structure group called the *holonomy group at  $u$* . It can be shown that, provided that  $M$  is connected, the holonomy groups at *all* bundle points  $u$  are isomorphic: the elements of the holonomy group at  $u'$  are conjugate to those of the holonomy group at  $u$ :  $g' = h_O g_O h^{-1}$  for some constant element  $h$  of  $G$ . The horizontal lift of  $C_p$  defines a *holonomy transformation*  $H(C_p)$ —an automorphism of the fiber  $F_p$  "above"  $p$ . Fixing some bundle point  $u$  "above" an arbitrary point  $p$  of  $M$ , we can now define the *holonomy of  $C_q$  at  $u$* :  $H_u(C_q)$  for a piecewise differentiable closed curve  $C_q$  beginning and ending at an arbitrary point  $q$  as the element of the holonomy group at  $u$  corresponding to a closed curve  $C'_p = \Gamma^{-1} * C_q * \Gamma$ , where  $\Gamma$  is a curve from  $p$  to  $q$ ,  $\Gamma^{-1}$  retraces the path from  $q$  to  $p$ , and  $*$  indicates composition of curves ( $H_u(C_q)$  is independent of which curve  $\Gamma$  is chosen). Then  $u_O H_u(C_q) = H(C'_p)[u]$ .

The holonomies will in general depend on the bundle connection. But distinct

connections may yield the same holonomy group, and so lead to essentially the same holonomies. One way to describe the relations between the holonomies of distinct connections is to express these connections in terms of their pull-backs onto  $M$  (where I assume for simplicity that the bundle admits a global section). Thus suppose that the representation of a connection on  $M$  in a given section is  $\mathbf{W}_\mu$ . Then the path-ordered operator

$$O(\exp[-i/\hbar \int_C W_\mu^i T_i \cdot dx^\mu])$$

defines an element  $g(C)$  of the structure group, where  $T_i$  are the generators of the structure group, and the  $W_\mu^i$  are the "internal" components of  $\mathbf{W}_\mu$ . Moreover, the action of this element defines a holonomy transformation of  $C$  at its starting point  $p$ :  $u_O g(C) = H(C)[u]$ , for  $u \in F_p$ .

Now consider two distinct bundle connections  $X^1, X^2$  represented in this way by  $\mathbf{W}_\mu^1, \mathbf{W}_\mu^2$ . The path-ordered operators for each closed curve  $C$  in  $M$  will yield group elements  $g_1(C), g_2(C)$  respectively. We now have the following result (Anandan(1980)).  $\mathbf{W}_\mu^1, \mathbf{W}_\mu^2$  are related by a gauge transformation corresponding to a vertical bundle automorphism if and only if, for every piecewise differentiable closed curve  $C$  in  $M$  starting at  $p$  there is a fixed element  $h$  of the structure group such that  $g_2(C) = h_O g_1(C)_O h^{-1}$ . ( $h$  establishes the isomorphism between the holonomy groups at the points  $u, v = u_O h$ .) What this implies is that two connections  $X^1, X^2$  are related by a vertical bundle automorphism if and only if, for every point  $p$  of  $M$ , there is a group element  $h$  such that for every bundle point  $v$  above  $p$ , and every piecewise differentiable closed curve  $C$  in  $M$ ,  $H_u^1(C) = H_v^2(C)$ , where  $v = u_O h$  and  $H_i$  represents a holonomy of  $X^i$  ( $i=1,2$ ). This gives the precise sense in which all and only connections related by a vertical

automorphism have the same holonomies: for each curve  $C$ , each such connection defines the same set of holonomies, relative to an arbitrary base point  $p$  of  $M$ .

Everything simplifies for an Abelian structure group, since in this case connections related by a vertical bundle automorphism agree on the group element corresponding to each curve  $C$  ( $g_1(C)=g_2(C)$ ), and this element may simply be taken as the holonomy for  $C$  ( $H^1(C)=H^2(C)$ ), independent of bundle point  $u$  or base point  $p$ . For example, the structure group for electromagnetism is the Abelian group  $U(1)$ , and the action of the holonomy for a closed curve  $C$  on the wave-function of charged particles just multiplies this by the Dirac phase factor for  $C$ .

According to a gauge theory, the observable phenomena in a gauge situation depend on the bundle connection only through its holonomies. So connections with the same holonomies are strongly empirically equivalent: even assuming the theory, no observation or experiment could help to discriminate between holonomy-equivalent connections. This motivates the view that it is the holonomies that represent the real physical structures in a gauge situation, rather than any particular bundle connection that is compatible with them. In capsule form, the view is that gauge potentials are holonomies.

What kind of object is a holonomy? Mathematically, it is an element of a Lie group--the symmetry group of a gauge theory. The fiber bundle formulation takes this as the structure group of the theory. But the view that gauge potentials are (faithfully represented by) holonomies can be stated and defended independently of a fiber bundle formulation.

Suppose one is given a differentiable map  $\varphi: L_p \rightarrow G$  from the set of piecewise differentiable closed curves in  $M$  starting at  $p$  into a symmetry group  $G$ , where  $\varphi$  takes the same value on curves whose images differ by the image of curves that enclose no area. Suppose further that  $\varphi$  is a homomorphism from the group of such curves (under composition) into  $G$ . Then Anandan(1983) shows that this defines a connection on a principal fiber bundle over  $M$  with structure group  $G$  with holonomy transformations  $H(C_p)$  determined by  $\varphi$  up to a constant similarity transformation  $g \in G: H'(C_p) = g_O H(C_p)_O g^{-1}$ . So one can take as basic a set of “holonomies” defined in terms of a homomorphism from closed curves on  $M$  into a symmetry group, and use this to construct a representation in terms of a connection on a principal fiber bundle. A fiber bundle formulation is thereby shown to provide merely a convenient and elegant mathematical expression rather than an essential foundation for the view that gauge potentials are faithfully represented by holonomies.

But what physical structures do holonomies represent? Call the image of each non-self-intersecting piece-wise differentiable closed curve  $C$  in  $M$  a *loop*. The holonomies of distinct curves with the same image are related by a similarity transformation. The set of such holonomies is a feature of the loop itself, invariantly represented by the so-called *Wilson loop*—the trace of every holonomy in the set. But the set of holonomies of a given loop itself has additional structure: these holonomies form components of an object with the same “internal” transformation properties as the associated gauge field. A gauge change on an open subset of  $M$  may be interpreted passively as reexpressing this object in a different set of “internal”

coordinates, or actively as transforming the object into a new object via a "gauge rotation". For example, the  $SO(3)$  holonomies of a loop have the structure of a vector in "internal" space, while the  $U(1)$  holonomy of a loop is a scalar. Thus the holonomies of a loop may be taken to represent a set of physical properties of that loop in just the way that any vector field or other mathematical object with well-defined transformation behavior represents a set of physical properties of an object at each point.

What is distinctive is not the properties represented by holonomies but the nature of the object whose properties they are. On the present view, holonomies represent global properties of a loop that are not determined by any intrinsic properties of the points on that loop. Call any process that involves intrinsic physical properties of or at a region of space-time that fail to supervene on intrinsic physical properties attached at space-time points in (or on the boundary of) that region *nonseparable*.<sup>7</sup> If the holonomies of a gauge theory faithfully represent global properties of space-time loops, then that theory describes nonseparable processes. For it ascribes intrinsic properties to space-time regions (the loops) that do not supervene on intrinsic physical properties of or at the space-time points that lie in these regions.

Even though holonomy properties of loops fail to supervene on intrinsic physical properties of or at the points that make them up, there is another supervenience thesis that does hold, at least in the absence of magnetic monopoles. To explain this thesis, note that some loops may be thought of as generated from others. Suppose loops  $L_1, L_2$  are the images of piecewise differentiable closed curves  $C_1, C_2$ , each beginning and ending at  $p$ . Define a closed

curve  $C=C_1*C_2$ , the composition of  $C_1,C_2$ , as a curve that first traces out  $C_2$ , then traces out  $C_1$ .

The holonomies of  $C$  are given by the group product of appropriate pairs of holonomies of  $C_1,C_2$ . Now  $L_1,L_2$  may contain overlapping segments. Define the sum  $L=L_1+L_2$  of  $L_1,L_2$  as the image of a closed curve  $C_+$  that traces out  $C_2$ , then  $C_1$ , but omits any segment that is traversed by  $C$  twice in opposite directions (see Figure 2). ~~PLACE FIGURE 2 ABOUT HERE~~ This may be extended naturally to finite sums. The holonomy properties of  $L=\sum_i L_i$  are wholly determined by those of the  $L_i$  that generate it: they are represented by the holonomies of the curve  $C$ , the composition of curves  $C_i$  that trace out the  $L_i$ . The holonomy properties of any loop are determined in this way by the holonomy properties of any set of loops that compose it. Moreover, in the absence of magnetic monopoles one can compose a given loop enclosing a surface out of tinier and tinier closed loops around points on that surface. In the limit, the holonomy properties of a finite loop are determined by those of any infinitesimal loops that compose it in this way. This gives what might be called the loop supervenience of holonomy properties.

Relative to a bundle section, the gauge potential may be represented on (a subset of)  $M$  by a Lie-algebra-valued one-form field. The associated gauge field is then represented by the corresponding curvature two-form. The curvature is related directly to the holonomies: each component of this curvature at a point  $p$  (where there is no magnetic monopole) relative to a coordinate system around  $p$  on  $M$  arises as the limit of the ratio of a holonomy around smaller and smaller closed curves enclosing  $p$  to the "area" (represented in terms of the curve

parametrization) enclosed by such a curve. The gauge field and the infinitesimal holonomies are then two sides of the same coin: each defines the other. In the absence of magnetic monopoles the loop supervenience of holonomy properties is equivalent to the supervenience of the holonomy properties of a loop on the gauge field properties at space-time points on a surface enclosed by the loop. For example, the electromagnetic properties represented by Dirac phase factors of a loop  $L$  are determined by the values of the electric and magnetic field strengths at points on a surface enclosed by  $L$ . Despite their global nature, holonomy properties supervene on local physical properties. But the holonomy properties of a finite loop remain global rather than local in the strong sense that some of the points whose local properties determine them are neither on nor even close to that loop.

**6. In Defense of Nonseparable Holonomies.** Ockham's razor cuts in favor of the view that gauge potentials are nonseparable holonomy properties. This view postulates no privileged, but empirically inaccessible, gauge. It postulates no infinity of ineffable objects associated with each space-time point. Holonomy properties, unlike novel gauge properties, may be coherently referred to and, moreover, empirically accessed through interference experiments. But there is a constraint on the operation of Ockham's razor--entities must not be multiplied *beyond necessity*. Is it necessary to acknowledge the existence of any of the additional local objects or properties postulated by alternatives to the nonseparable holonomy properties view?

Arguments in favor of the existence of some such additional structure may be either

metaphysical or epistemological. One metaphysical argument takes the form of an objection to the nonseparable holonomy properties view. The objection is that holonomy properties cannot be nonseparable, since it is metaphysically unacceptable to suppose that space-time regions have global properties not determined by local properties at their constituent points. No scientific theory may postulate such "free-floating" global properties. If a gauge theory is to remain within the realm of science, rather than postulating miracles, it must provide some underlying local structure--novel gauge properties or objects--to provide the necessary metaphysical support for holonomy properties.

It seems to me that there is remarkably little to this metaphysical objection. It is reminiscent of Cartesian objections to Newtonian science. The response is straightforward: Nonseparable holonomy properties are no more "occult" or unscientific than Newtonian forces. In both cases, empirical success of the theory that postulates the allegedly problematic structures justifies belief in them, independently of whether or not further theorizing will turn out to provide a more familiar grounding for these structures.

Epistemological objections may seem more promising. The idea here is that if the gauge theory is true, then there are striking facts that can satisfactorily be explained only by the postulation of local theoretical structure that determines holonomy properties. Inference to the best explanation therefore warrants belief that such structure indeed underlies these global properties.

Two kinds of facts come to mind in this context.<sup>8</sup> The first concerns relations among



holonomy properties of loops. Suppose that loop  $L$  is generated by loops  $L_1, L_2$ : then the holonomy properties of  $L$  may be represented by a curve  $C$  composed of the curves  $C_1, C_2$  that trace out  $L_1, L_2$ . Loosely, the holonomy properties of  $L$  are just the products of those of  $L_1, L_2$ . One can explain this relationship on the novel gauge properties view by supposing that, for each open path  $\Gamma$  on  $M$ , the connection on a principal bundle over  $M$  represents some intrinsic relation among intrinsic gauge properties of points along  $\Gamma$ . But if there is no such relation and no such properties, then the relationship appears as a puzzling, brute fact.

It is not clear how good an explanation this would be, given the epistemic inaccessibility of the additionally postulated relation and properties. To see its failings, compare the Newtonian explanation of why all inertially moving objects move at constant velocities with respect to one another: Inertial motion corresponds to moving through equal intervals of absolute space in equal intervals of absolute time. This is not a good explanation, precisely because nothing in Newton's theory (or elsewhere) enables experiments to yield information relevant to determining the state of absolute rest. A structure, like absolute rest, that remains radically indeterminate, even given a formulation of a theory that postulates it, does not provide the basis for a credible explanation of a generalization in the domain of that theory. A good theoretical explanation of such a generalization has unifying power--power that accrues when a postulated theoretical structure plays other roles in this or other theories. And typically this is how that structure becomes theoretically determinate. (Consider how the diverse explanatory roles played by Avogadro's number permitted multiple independent determinations of its value,

while demonstrating the explanatory power of the principle that a mole of gas always contains the same number of molecules.) I conclude that even a firm believer in inference to the best explanation in theoretical science may reasonably reject an inference to additional structure in a gauge theory beyond the holonomy properties themselves.

Moreover, the relations among holonomy properties can be given an alternative explanation which appeals to no such additional structure. For these follow simply from loop supervenience, or, equivalently in the absence of magnetic monopoles, from the (global, not local) determination of holonomy properties by gauge field strengths. These explanations appeal to no empirically inaccessible or indeterminate structures. In their light the relations among holonomy properties no longer appear as disconnected brute facts that demand explanation in terms of an additional level of "hidden" structure.

Leeds(1999) has argued that only by adopting an interpretation of EM theory according to which the vector potential is faithfully represented by a particular connection on a principal fiber bundle can one explain why the canonical momentum operator takes the form  $-i\hbar\nabla - q\mathbf{A}$  for a particle with charge  $q$  subject to a magnetic vector potential  $\mathbf{A}$ .<sup>9</sup> He motivates his view by proposing that an adequate interpretation of electromagnetism must take phases seriously. This means taking seriously some 'element of reality' that shows up in the various gauges as the phase factor along open as well as closed paths. In the context of a fiber bundle formulation, "taking phases seriously" means accepting that electromagnetism is uniquely represented by some privileged connection on a principal bundle.

There is a three-part reply to Leeds's argument. Note first that a momentum operator is usually<sup>9</sup> *called* canonical just when quantization of a classical theory in its Hamiltonian form converts the classical canonical momentum quantity in that theory into this operator. But the classical quantity that plays the role of canonical momentum in a Hamiltonian formulation of classical electromagnetism is just  $\mathbf{p}-q\mathbf{A}$ , which becomes  $-i\hbar\nabla-q\mathbf{A}$  on quantization. Lest this be dismissed as a cheap semantic trick, note secondly that only gauge-invariant quantities are candidates for observable quantities after canonical quantization, and  $-i\hbar\nabla-q\mathbf{A}$  is indeed gauge-invariant, unlike  $-i\hbar\nabla$ . Now canonical quantization is not always a well-defined procedure, and is not in any case guaranteed to yield a correct theory. So the fact that  $-i\hbar\nabla-q\mathbf{A}$  is a natural gauge-invariant candidate for representing the physical quantity momentum does not conclusively prove that this must be the correct form for the momentum operator in an empirically adequate theory of electromagnetic interactions of quantized particles. But we have techniques for measuring momentum that yield results in the absence of electromagnetism whose values and statistics conform to quantum mechanical predictions. Applying those same techniques in the presence of electromagnetism yields (gauge-independent) results whose values and statistics conform to quantum mechanical predictions only if momentum is there represented by the operator  $-i\hbar\nabla-q\mathbf{A}$ . Ultimately, it is only experiment that decides that this operator represents the quantity whose values our momentum-measurements give us.

To defend the nonseparable holonomy properties interpretation of a gauge theory I contrasted it favorably with the novel gauge properties view. This involved the claim that the

novel gauge properties are problematic because they permit one to formulate unanswerable questions. But an advocate of the novel properties view could deny that this would render these properties problematic.<sup>10</sup> It may be an epistemological defect in a theory to raise meaningful but unanswerable empirical questions. But the complaint against the novel gauge properties view is rather that there are questions it doesn't even permit one meaningfully to ask. And that is not obviously an epistemological defect in a theory. Indeed, it might be touted as a virtue of a theory that it renders metaphysical questions not merely empirically unanswerable but literally meaningless!

However, in the present situation the onus is on a proponent of the novel gauge properties view to persuade the skeptic that accepting a gauge theory commits him or her to the existence of novel gauge properties. Such properties are not observed in any experiment, and nor are they mentioned in any standard formulation of a gauge theory. The first step in arguing for their existence must surely be to provide a coherent way of talking about them. If we cannot use language to pick out distinct novel properties then we will be unable to make the kind of meaningful assertion about them that could lead to testable predictions or credible explanations of known phenomena. A view that alleges the existence of such properties but does not permit such meaningful assertions about them does not constitute a rival interpretation of a gauge theory as a scientific theory. Instead, it supplements the nonseparable holonomy properties interpretation with a metaphysical overlay which, though perhaps satisfying to some, is no part of the physics itself.

## REFERENCES

Aharonov, Yakir and David Bohm (1959), "Significance of Electromagnetic Potentials in the Quantum Theory", *Physical Review* 115: 485-91.

Aharonov, Yakir and J.L. Safko (1975), "Measurement of Noncanonical Variables", *Annals of Physics* 91: 279-294.

Anandan, Jeeva (1980), "Quantum Interference and the Classical Limit", *International Journal of Theoretical Physics* 19: 537-56.

———(1983), "Holonomy Groups in Gravity and Gauge Fields", in Denardo, G. and H.D. Doebner (eds.), *Proceedings of the Conference on Differential Geometric Methods in Theoretical Physics, Trieste (30 June to July 1981)* (Singapore: World Scientific),

Feynman, Richard P., R.B. Leighton, and M.L. Sands (1965), *The Feynman Lectures on Physics, Volume 2* (Reading, Mass.: Addison Wesley).

Fine, Arthur and Jarrett Leplin (eds.) (1989), *PSA 1988 Volume Two* (East Lansing, Michigan: Philosophy of Science Association).

Gribov, V.N. (1977), "Instability of Non-Abelian Gauge Theories and Impossibility of Choice of Coulomb Gauge", Translated from a lecture at the 12<sup>th</sup> Winter School of the Leningrad Nuclear Physics Institute, SLAC-TRANS-0176.

Healey, Richard (1991), "Holism and Nonseparability", *Journal of Philosophy*, LXXXVIII: 393-421.

———(1994), "Nonseparable Processes and Causal Explanation", *Studies in History and Philosophy of Science* 25: 337-374.

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

———(1999), "Quantum Analogies: a Reply to Maudlin", *Philosophy of Science* 66: 440-447.

Hong-Mo, Chan and Tsou, Sheung Tsun (1993) *Some Elementary Gauge Theory Concepts* (Singapore: World Scientific).

Kobayashi, S. and Nomizu, K. (1963) *Foundations of Differential Geometry, Volume 1* (New York: Interscience).

Leeds, Steven (1999), "Gauges: Aharonov, Bohm, Yang, Healey", *Philosophy of Science* 66: 606-627.

Maudlin, Tim (1998), "Discussion: Healey on the Aharonov-Bohm Effect", *Philosophy of Science* 65: 361-368.

Singer, I.M.(1978), "Some Remarks on the Gribov Ambiguity", *Communications in Mathematical Physics* 60: 7-12.

Wu, T.T. and Yang, C.N. (1975), "Concept of Nonintegrable Phase Factors and Global Formulation of Gauge Fields", *Physical Review D* 12: 3845-57.

## FOOTNOTES

1. This and subsequent section headings neither advocate nor presuppose an uncritical scientific realism. What is at issue is the physical content of certain gauge theories, not their truth, and certainly not the general character of the relation of a scientific theory to a noumenal world or to its creators and users.
2. As I pointed out in my(1999) in response to Maudlin 1998, who offered the present suggestion, while characterizing it as "simplistic".
3. If the position operator had a discrete spectrum one could think of each of these representations as simply corresponding to a choice of a different eigenvector from each one-dimensional eigenspace of the position operator, with no change in the eigenspaces themselves.
4. For which, see the papers by Norton, Butterfield and Maudlin in Fine et al. 1988, 56-91.
5. I am indebted to Tim Maudlin for a spirited defense of this proposal.
6. As Singer(1978) explains, a map that selects a particular representative from each gauge-equivalent class of connections on a bundle with a non-Abelian structure



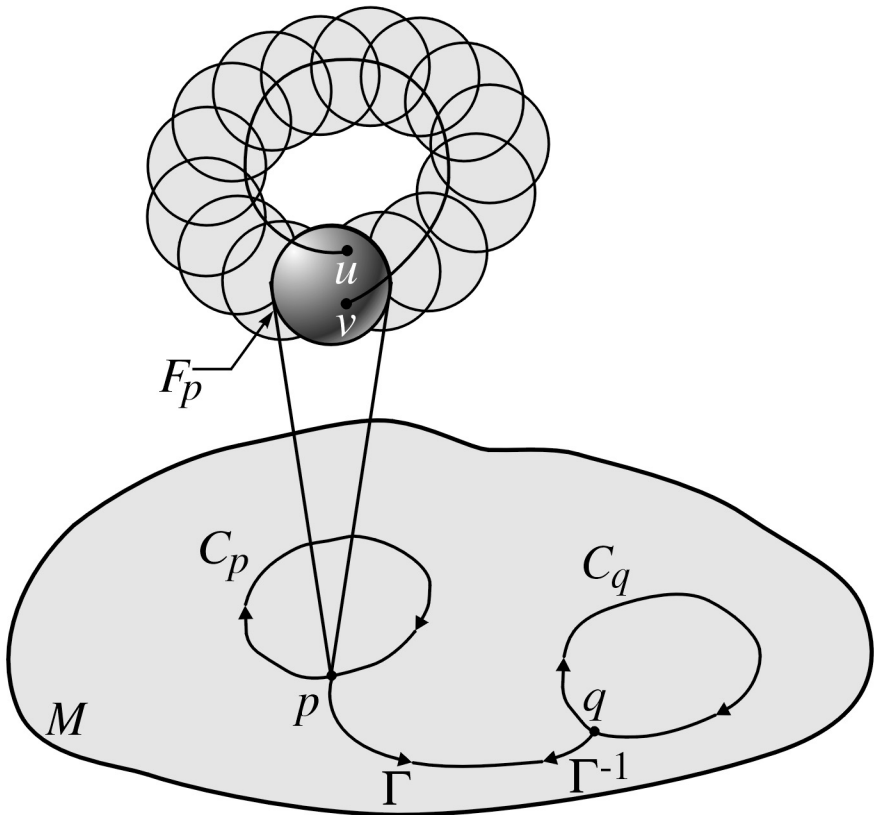
group like SU(2) cannot, in general, be continuous. This generalizes Gribov's 1977 result that there is no simple analogue to the Coulomb gauge for non-Abelian gauge theories.

7. Intuitively, a property is intrinsic to an object just in case its possession by that object is independent of the existence or properties of anything else. Philosophers have found it hard to turn this intuition into anything more precise. For further discussion of both this issue and the concept of nonseparability, see Healey 1991,1994.

8. Here I am indebted to Leeds 1999 and Maudlin in correspondence.

9. Leeds gives the form of the canonical momentum operator as  $\hbar/i \nabla - eA/c$ . I use  $q$  instead of  $e$  for the particles' charge and units in which  $c=1$ . Others have used the term 'canonical' to refer instead to a momentum operator  $\mathbf{p}$  of the form  $\mathbf{p} = -i\hbar \nabla$ , following Aharonov and Safko 1975.

10. I am much indebted to Tim Maudlin for pressing this objection in a lively correspondence.



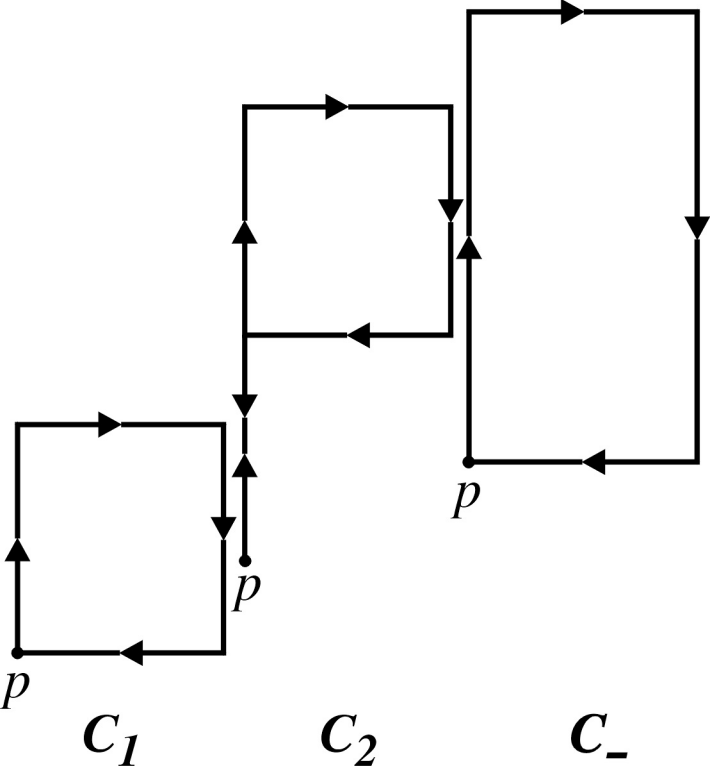


Figure 1. The curve from  $u$  to  $v$  is the horizontal lift of  $C_p$ . The curve  $C'_p$  results from composing the curves  $\Gamma$ ,  $C_q$  and  $\Gamma^{-1}$ .

Figure 2. The curves  $C_1$ ,  $C_2$ ,  $C_-$  all begin and end at the same point  $p$ .  $C_-$  traces out the loop  $L = L_1 + L_2$ , where  $C_1$  traces out the loop  $L_1$  and  $C_2$  traces out the loop  $L_2$ .