The Principles of Gauging*

Holger Lyre†‡

Ruhr-University Bochum

The aim of this paper is twofold: First, to present an examination of the principles underlying gauge field theories. I shall argue that there are two principles directly connected to the two well-known theorems of Emmy Noether concerning global and local symmetries of the free matter-field Lagrangian, in the following referred to as "conservation principle" and "gauge principle". Since both these express nothing but certain symmetry features of the free field theory, they are not sufficient to derive a true interaction coupling to a new gauge field. For this purpose it is necessary to advocate a third, truly empirical principle which may be understood as a generalization of the equivalence principle. The second task of the paper is to deal with the ontological question concerning the reality status of gauge potentials in the light of the proposed logical structure of gauge theories. A nonlocal interpretation of topological effects in gauge theories and, thus, the non-reality of gauge potentials in accordance with the generalized equivalence principle will be favoured.

1. The Gauge Argument. Textbook presentations of the logical structure of gauge field theories usually emphasize the importance of the *gauge principle* (cf. Aitchison and Hey 1982, p. 176): Start with a certain free field theory—take Dirac's theory $\mathcal{L}_D = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$ for instance—and consider local gauge transformations

$$\psi(x) \to \psi'(x) = e^{iq\alpha(x)}\psi(x). \tag{1}$$

To satisfy the requirement of local gauge covariance of $\mathcal{L}_D$ the usual derivative has to be replaced by a covariant derivative

$$\partial_\mu \;\to\; D_\mu = \partial_\mu - iqA_\mu. \tag{2}$$

Thus, instead of a free field theory, we obtain—in a somewhat miraculous way—a theory with interaction

$$\mathcal{L}_D \;\to\; \mathcal{L}'_D = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + q\bar{\psi}\gamma^\mu A_\mu\psi. \tag{3}$$

Besides this, in usual textbook presentations reference is also made to the importance of "Noether's theorem": The existence of a global (i.e. rigid) k-dimensional symmetry group is connected with the existence of k conserved currents. This general result may very well be referred to as a principle of its own, here called the *conservation principle*. In the case of Dirac's

*Talk at PSA2K meeting in Vancouver, B.C., November 4-7, 2000. To be published in *Philosophy of Science*.

†Institut für Philosophie, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Email: holger.lyre@ruhr-uni-bochum.de

‡Special thanks to Tim Oliver Eynck for helpful remarks.

theory we find that $\mathcal{L}_D$ exhibits global gauge covariance under $\psi(x) \to \psi'(x) = e^{iq\alpha}\psi(x)$. The Noether current then reads $j^\mu(x) = -q\bar{\psi}(x)\gamma^\mu\psi(x)$. Now, the "miracle" of the gauge principle consists in the idea that by simply postulating local gauge covariance one is led to introduce a new interaction potential $A_\mu(x)$ obeying $A_\mu(x) \to A'_\mu(x) = A_\mu(x) - \partial_\mu\alpha(x)$. Surely, it is very tempting to hold this suggestive interpretation, but are we really forced to?

From a mathematically more rigorous point of view, the above presentation is a bit to much of a "miracle". Both the conservation principle as well as the gauge principle are simply concerned with Noether's first and second theorem–and thus with a mere symmetry analysis of the free field theory. Let $\phi_i(x)$ be a field variable and let i be the index of the field components, then Noether's first theorem states that the invariance of the action functional $S[\phi] = \int \mathcal{L}[\phi_i(x), \partial_\mu\phi_i(x)]\, d^4x$ under the action of a k-dimensional Lie group implies the existence of k conserved currents. This is what is usually just called "Noether's theorem". In the language of the underlying fiber bundle structure, Noether's first theorem points to the importance of the bundle structure group—in the case of the above Dirac-Maxwell theory the gauge group $G = U(1)$. Hence, we are working with a $U(1)$-principal bundle $\mathbb{P}$ over spacetime.[1]

Now let $\mathcal{G} = \mathcal{A}ut(M) \simeq \mathcal{D}iff(M) \circledS G$ be the automorphism group of $\mathbb{P}$. Noether's second theorem, then, states that the invariance of the action $S[\phi]$ under $\mathcal{G}$ implies the existence of k constraints known as Bianchi identities. From (2) we first of all find the Jacobi identity $\epsilon^{\mu\nu\rho\sigma}[D_\nu, [D_\rho, D_\sigma]] = 0$. With the definition $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ this is equivalent to the Bianchi identity $\epsilon^{\mu\nu\rho\sigma} D_\nu F_{\rho\sigma} = 0$. Clearly, $\mathcal{G}$ is infinite-dimensional and consists of spacetime-dependent (i.e. "local") group elements. In this way Noether's second theorem gives rise to the postulate of local gauge covariance which in turn underlies the gauge principle.

Note again that Noether's analysis is just concerned with the symmetry conditions of a given action functional. Hence, this does not allow for—or even force us— to interpret the A_μ-term as a new interaction term. Both the conservation as well as the celebrated gauge principle lay claim to certain symmetry conditions of a given field theory—without introducing a new field. Indeed, how could a new physical field be derived from a mere analysis of the symmetry structure of some theory? How, then, are we to understand the occurence of the A_μ-term?

2. Intrinsic Gauge Theoretic Conventionalism. In recent times, a critical reading of the gauge argument has been presented by several authors; cf. Brown (1999), Healey (2000), Teller (2000). To easily see the issue consider a wavefunction $\Psi(x) = \langle x|\phi\rangle$ in the position representation $|x\rangle$ (with $\left\{|\phi\rangle\right\}$ spanning an abstract Hilbert space). Now, local gauge transformations read $|x\rangle \to |x'\rangle = e^{i\alpha(x)}|x\rangle = \hat{U}|x\rangle$. Such a transformation acts as changing the representation basis of the Hilbert space and, thus, operators on that Hilbert space have to be transformed, too. A general operator transformation looks like $\hat{O}' = \hat{U}\hat{O}\hat{U}^+$. In the particular case of the derivative operator we find $\partial_\mu \to D_\mu = \partial_\mu - iqA_\mu(x)$ with the definition $A_\mu(x) = -\partial_\mu\alpha(x)$.

We therefore clearly see that (2) has to be understood as a mere change in the position representation expressed in terms of local gauge transformations. Hence, the covariant repre-

[1] A more detailed presentation of gauge theories and their bundle structure as well as a brief account of the theory of fiber bundles can be found in Guttmann and Lyre (2000).

sentation of the derivative is as conventional as a mere coordinate representation. This feature might very well be called an "intrinsic gauge theoretic conventionalism". The clear consequence of this is that the celebrated gauge principle is not sufficient to derive the coupling to a new interaction-field. No new physics enters, no new physical field is really introduced!

3. A Missing Principle. A true gauge field theory should be considered as a coupling between a matter-field and an interaction-field theory.[2] We are therefore faced with the following problem. We have, on the one hand, equations of motion of the free matter-field (e.g. Dirac's equation). Due to the gauge principle the Lagrangian reads

$$\mathcal{L}'_D = \mathcal{L}_D + \mathcal{L}^{(i)}_{inhom} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \jmath^{(i)}_\mu A^\mu \tag{4}$$

with an unphysical inhomogeneity term, since the connection field A^μ is flat (i.e. the curvature gauge field vanishes). On the other hand we have certain gauge field equations (Maxwell or Yang-Mills equations)

$$\mathcal{L}'_F = \mathcal{L}_F + \mathcal{L}^{(f)}_{inhom} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \jmath^{(f)}_\mu A^\mu. \tag{5}$$

Here, the inhomogeneity stems from the field sources, i.e. certain "field charges" $q^{(f)}$. In contrast to this the vector current $\jmath^{(i)}$ in (4) implies a factor $q^{(i)}$ which is due to the phase $e^{iq^{(i)}\alpha}$ of the Dirac wavefunction. As will be explained in a moment we will call this the "inertial charge".

In (4), the inhomogeneity term $\mathcal{L}^{(i)}_{inhom}$ stems from the gauge principle. It should be clear from our critical reading of this principle that there is no *a priori* possibility to identify $\mathcal{L}^{(i)}_{inhom}$ and $\mathcal{L}^{(f)}_{inhom}$, or $q^{(i)}$ and $q^{(f)}$, respectively. Since both conservation principle as well as gauge principle turned out as mere analytic statements about the symmetry structure of the free matter-field theory, we are in need of a truly empirical—synthetic so to speak—principle of gauging, which allows for the identification of $\mathcal{L}^{(i)}_{inhom}$ and $\mathcal{L}^{(f)}_{inhom}$. Fortunately, the gauge theoretic analogy to general relativity may help to find such a missing principle.

In fact, in standard general relativity we may also formulate a gravitational gauge principle. The starting point for this is the free geodesic equation $\frac{d}{d\tau}\theta^\mu_\alpha(\tau) = 0$ for a tetrad reference frame θ^μ_α. The gauge principle demands covariance under local $SO(1,3)$ or $\mathbb{R}^{(1,3)}$ transformations.[3] We get a covariant derivative

$$\frac{d}{d\tau}\theta^\mu_\alpha(\tau) \;\rightarrow\; \nabla_\tau \theta^\mu_\alpha(\tau) = \frac{d}{d\tau}\theta^\mu_\alpha(\tau) + \Gamma^\beta_{\nu\alpha}\frac{dx^\nu(\tau)}{d\tau}\theta^\mu_\beta(\tau). \tag{6}$$

Now, the Levi-Civita connection Γ_μ does not necessarily represent a true gravitational potential with a non-vanishing gravitational field (i.e. Riemann curvature). Indeed, Γ_μ might occur simply because of a peculiar choice of coordinates!

---

[2]For an elaboration of the next two sections the reader may want to refer to my companion paper Lyre (2001).

[3]It depends on the Noether current arising from the conservation principle how to couple the gravitational field. Certainly, a straightforward choice for the gauge group of general relativity is the group of Poincaré translations $\mathbb{R}^{(1,3)}$ which implies the conservation of energy-matter. Moreover, this is a reasonable choice since local translations are equivalent to diffeomorphisms.

The "true" gravitational field with non-vanishing Riemann curvature is of course governed by the Einstein field equations $R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = -\kappa\, T_{\mu\nu}$. The r.h.s. represents the field source, i.e. gravitational mass $m^{(g)}$ which is encoded in the energy-momentum tensor $T_{\mu\nu}$. In contrast to this, a freely moving observer in spacetime (a reference frame represented by a tetrad in the geodesic equation) will be assigned an inertial mass $m^{(i)}$. As the reader might guess already, in general relativity the conceptual problem of linking equations of motion and field equations is solved on the basis of the *equivalence principle*. The cruicial identification

$$m^{(i)} = m^{(g)} \tag{7}$$

becomes indeed the one and decisive empirical input to any geometric theory of gravitation. There is no *a priori* reason to identify inertial and gravitational mass and, hence, the equivalence principle has to be vindicated by the experimental fact that different materials do have the same free fall behaviour. In this way, the universality of the gravitational coupling constitutes a deep fundamental insight.

4. The Generalized Equivalence Principle. We may now return to our problem of finding the missing empirical principle in gauge field theories. The relativistic equivalence principle implies the possibility of a non-flat connection and, hence, a non-vanishing gravitational field. Due to the close analogy between general relativity and standard quantum gauge field theories we may very well *generalize* the idea of the equivalence principle. Indeed, the equivalence of inertial and field charges

$$q^{(i)} = q^{(f)} \tag{8}$$

solves our problem and allows for a true coupling term $\mathcal{L}_{coup} = \mathcal{L}^{(i)}_{inhom} = \mathcal{L}^{(f)}_{inhom}$. Thus, equations of motion and field equations belong to one combined framework and we obtain the full Lagrangian of a gauge field theory representing—quite generally—the coupling between matter-field and interaction-field

$$\mathcal{L}_{GFT} = \mathcal{L}_D + \mathcal{L}_{coup} + \mathcal{L}_F. \tag{9}$$

We may give the following geometric formulation of the *generalized equivalence principle*:

> **GEP:** *It is always possible to perform a local gauge transfomation such that, locally (i.e. at a point), the gauge field vanishes.*

In this way, GEP implies a non-flat connection, i.e. a gauge potential which is irrevocably connected with the occurrence of an interacting gauge field originating in the field charges and obeying its own dynamics. Equation (8) turns out as a direct consequence of this, for if we regard the connection as non-flat, the field must have its sources in certain field charges. Moreover, GEP includes the interaction-free theory as a local limiting case.

The reader may wonder whether we have simply replaced one miracle by another. However, the equivalence (8) is far from trivial. There is—quite analogous to (7)—no *a priori* reason to identify inertial and field charges. Let us assume for a moment $\frac{q^{(f)}}{q^{(i)}} \neq 1$. This means that different types of particles of equal electric charge would couple differently to the electromagnetic

field. We should expect a difference in the coupling of electrons and muons or d-quarks and s-quarks—to give but two examples—and should therefore write down different Dirac equations

$$(i\gamma^\mu\partial_\mu - m)\psi e^{iq^{(i)}\alpha} = c_p\ q^{(f)}\ \gamma^\mu A_\mu\ \psi e^{iq^{(i)}\alpha} \tag{10}$$

for different types of particles with the same $q^{(f)}$ but a particle-type dependent factor c_p. This is clearly not what we observe. In fact, GEP predicts a whole variety of *null-experiments* (as does its relativistic counterpart). The equivalence (8) indicates the empirically known universality of the gauge field coupling, turning GEP into the one and decisive physical principle of gauge theories.

The three principles of gauging We may summarize our considerations so far. It will be helpful to draw the following distinction of types of fiber bundles occuring in gauge theories: We may have trivial bundles[4] with flat and non-flat connections. Let us call them type 1 and type 2 bundles. As long as we are concerned with trivial bundles, the notion of a fiber bundle is in a way superfluous (since we may simply use a direct product). For non-trivial bundles, however, the fiber bundle framework becomes indispensible. Let us indicate non-trivial bundles as type 3 bundles. Since we may again distinguish between flat and non-flat connections, we may accordingly call them type 3a and type 3b bundles.

I shall rewiew the three proposed principles of gauging:

Conservation principle. Based on Noether's first theorem it connects the global (i.e. rigid) symmetry of a free field theory with the existence of certain conserved quantities. As an analytic statement of the symmetry structure of the theory it does not contain any new physical information.

Gauge principle. Based on Noether's second theorem it connects the local (i.e. spacetime-dependent) symmetry of a free field theory with the *suggested* structure of the coupling to an interaction-field. It is tempting to take the suggested coupling already for granted, however, the gauge principle only implies flat connections and, hence, no non-vanishing interaction fields. In other words, the gauge principle does not allow for a transition from type 1 to type 2 bundles (or type 3a to 3b, respectively). As a mere analytic statement of the symmetry structure of the theory it also does not contain new empirical information.

Equivalence principle. This is a true empirical principle which lays claim for the universality of the gauge field coupling due to the identification $q^{(i)} = q^{(f)}$. This manifests the coupling of matter-fields and interaction-fields and allows to combine equations of motion and field equations into one framework. The equivalence principle implies the existence of non-flat connections and therefore non-vanishing interaction-fields. It is a synthetic statement of the empirical basis of gauge field theories.

A schematic representation is given in figure 1.

[4]Trivial bundles allow for global sections, they globally look like direct products spaces. In contrast to this, non-trivial bundles only locally look like a direct product.

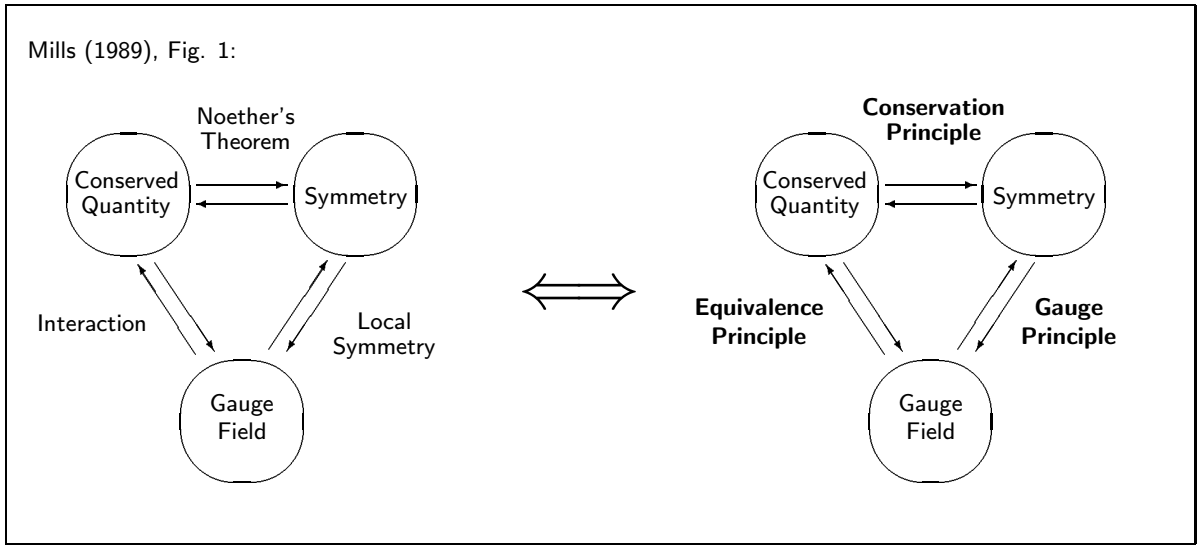

Figure 1: In his 1989 review article, Robert Mills gave a graphical representation of what he calls the "logical pattern of a gauge theory". The triangular structure of his pattern, on the left hand side, also illustrates our understanding of the principles of gauging in terms of the right hand side figure.

6. The Reality of Gauge Potentials. "Only gauge-independent quantities are observable." This truism is supported by our critical remarks on the intrinsic gauge theoretic conventionalism of local symmetries. It is also in accordance with GEP as an argument in favour of non-flat connections, i.e. non-vanishing gauge-independent field strengths. Therefore, GEP lays claim for not considering gauge potentials as physically real entities. Clearly, this is true for type 2 and type 3b bundles which are concerned with non-flat connections. However bundles of type 3a seem to allow for physical—viz. topological—effects which have their origin in flat connections (type 1 is just a trivial case). Does this contradict GEP's point of view of not considering gauge potentials as physically real?

Usually physicists think along these lines. They do consider gauge potentials as real entities because of topological effects in field theories. This view is supported by some kind of a common-sense indispensability argument: First, gauge potentials—and matter-fields—are the genuine objects in the fiber bundle formulation of gauge theories. They are clearly indespensible for the mathematical formulation (as being the connection forms). Also, they are indespensible for the physical formulation of quantum field theories, since both the *coupling structure* (vertex structure) as well as the *quantization procedure* itself are represented on the level of potentials and not the field strengths. How, then, are we to do physics without potentials?

As Michael Redhead (2000) has pointed out, the situation is even worse, since no matter wether we consider potentials real or not, we will always face ontological problems. Quite generally, such problems seem to arise in theories with a certain mathematical *surplus structure*. Here

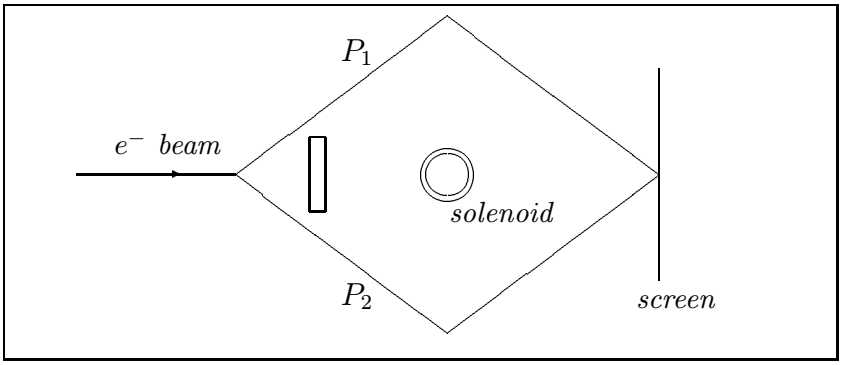

Figure 2: Schematic experimental configuration of the AB effect.

we have a mathematical structure M' which is larger than the structure M needed for a direct correspondence (i.e. isomorphism) to the observable physical structure P. The complement of M in M' might be called surplus structure. Gauge potentials are an example of surplus structure in gauge theories. Now, "Redhead's dilemma" looks like the following: On the one hand the reality of gauge-dependent potentials implies a mystic influence from non-observable physical beables to observable ones. This, in fact, is a version of the famous hole argument—and this first horn of the dilemma leaves us with indeterminism.[5] The second horn is that once we assert the non-reality of gauge potentials, this implies a "Platonist" role for mathematical elements to influence physical beables.

To get along with Redhead's dilemma we shall take a closer look to the well-known Aharonov-Bohm effect, which is indeed the paradigm case of type 3a bundles—and, hence, topological effects. Due to Aharonov and Bohm (1959) a shift in the interference pattern of the electron wave function surrounding a solenoid (on paths P_1 and P_2) is observed, even though the electron is shielded from the region of the magnetic field (see figure 2). Since only the vector potential has a non-vanishing contribution outside the solenoid, the AB effect is usually understood as showing the physical significance of gauge potentials.

As can be seen from the experimental configuration, the existence of the AB effect depends crucially on the fact that the configuration space of the electron is not simply-connected. Since the electron is shielded from the solenoid, this space has essentially the toplogy of a circle (as represented by any closed loop surrounding the solenoid, such as paths P_1 and P_2, for instance). Now, as Yang (1974) has first pointed out, the AB effect may be described solely in terms of the Dirac phase factor

$$\Delta\alpha = \oint_C A_\mu \, dx^\mu. \tag{11}$$

This integral lives in the space of loops and is called a holonomy. Clearly, holonomies are gauge-independent quantities and therefore appropriate candidates of observable entities.

So far, this does not solve our problem since we are still working with an integral which

[5]For a discussion of the bundle space hole argument see Lyre (1999).

depends on the gauge potential (in a gauge-independent manner, though). However, due to Stokes' theorem

$$\oint_{\mathcal{C}} A_\mu \, dx^\mu = \int_{\mathcal{S}} F_{\mu\nu} \, ds^{\mu\nu}, \tag{12}$$

we might very well describe the AB effect as a *nonlocal effect* in terms of the magnetic field strength alone. In fact, Stokes' formula allows to shift back and forth between the potential and the field strength interpretation. Presented this way, the AB effect turns out as a nice case study of theory underdetermination by empirical evidence. Physicists tend to favour the potential interpretation since it apparently allows for a local interaction account. This, however, leaves Redhead's dilemma unsolved.

A second, even stronger worry against the physicist's common line of simply accepting the reality of gauge potentials, is the fact that, in any case, due to the topological origin of Dirac's phase factor, we will never completely get rid of a certain kind of non-locality—or non-separability (Healey 1997). Topological effects unavoidably lead, in one way or the other, to a nonlocal account. It is therefore impossible to give a purely local description of the interference shift, neither within the field nor the potential interpretation. We may gladly accept the field interpretation and, also, should consider holonomies as physically real. This option has the clear advantage of avoiding Redhead's dilemma, since no surplus structure arises.

7. Conclusion. Even for the description of toplogical effects in type 3a bundles, the reality of gauge potentials is not enforced. We may very well represent the physically significant structures in an ontological universe consisting of matter-fields, gauge field strengths and holonomies. The price we pay is to accept a certain type of nonlocality in gauge theories—which seemingly differs from quantum nonlocalities such as EPR correlations due to its manifest topological origin, but seems unavoidable anyhow in both the potential and the field strength account.

Now, since holonomies may be represented in terms of gauge field strengths (because of Stoke's formula), we are in perfect agreement with GEP as an argument against the significance of flat connections. Indeed, the three proposed principles of gauging prove to be a consistent framework of the main conceptual structure of gauge field theories. Maybe, therefore, the idea of a generalized equivalence principle helps to clarify the issue of the logical gauge theoretic pattern. Nevertheless, a couple of deep philosophical puzzles remain to be solved—last but not least the very idea of gauging itself, which may heavily lean on a sufficient account of locality and nonlocality in physics. Thus, the issue of gauge theories should become much more the focus of philosophers of science than it was before. Personally, I couldn't agree more to how Michael Redhead (2001) has recently put it: *"The gauge principle is generally regarded as the most fundamental cornerstone of modern theoretical physics. In my view its elucidation is the most pressing problem in current philosophy of physics."*

References

Aharonov, Yakir and David Bohm (1959), Significance of electromagnetic potentials in the quantum theory. *Physical Review* 115(3): 485–491.

Aitchison, Ian J. R. and Anthony J. G. Hey (1982), *Gauge Theories in Particle Physics - A Practical Introduction.* Bristol: Hilger.

Brown, Harvey R. (1999), Aspects of objectivity in quantum mechanics. In Jeremy Butterfield and Constantine Pagonis (eds.), *From Physics to Philosophy.* Cambridge: Cambridge University Press.

Guttmann, Yair M. and Holger Lyre (2000), Fiber Bundle Gauge Theories and "Field's Dilemma". E-print arXiv:physics/0005051.

Healey, Richard (1997), Nonlocality and the Aharonov-Bohm effect. *Philosophy of Science* 64: 18–41.

Healey, Richard (2000), On the reality of gauge potentials. Preprint.

Lyre, Holger (1999), Gauges, Holes, and their 'Connections'. To appear in Don A. Howard (ed.), Proceedings of the "Fifth International Conference on the History and Foundations of General Relativity", Notre Dame, Indiana. (E-print arXiv:gr-qc/9904036).

Lyre, Holger (2001), A generalized equivalence principle. International Journal of Modern Physics D 10, in print. (E-print arXiv:gr-qc/0004054).

Mills, Robert (1989), Gauge fields. *American Journal of Physics* 57(6): 493–507.

Redhead, Michael (2000), The intelligibility of the universe. Forthcoming in A. O'Hear (ed.), Philosophy at the New Millennium.

Redhead, Michael (2001), The interpretation of gauge symmetry. To appear in Meinard Kuhlmann, Holger Lyre and Andrew Wayne (eds.), Proceedings of the International Conference on "Ontological Aspects of Quantum Field Theory", October 11-13, 1999, Bielefeld, Germany.

Teller, Paul (2000), The gauge argument. *Philosophy of Science* 67 (Supplement): S466–S481. (PSA98 Proceedings, ed. by Don A. Howard).

Yang, Chen Ning (1974), Integral formalism for gauge fields. *Physical Review Letters* 33(7): 445–447.