Gauge Matters

John Earman
University of Pittsburgh

The constrained Hamiltonian formalism is recommended as a means for getting a grip on the concepts of gauge and gauge transformation. This formalism makes it clear how the gauge concept is relevant to understanding Newtonian and classical relativistic theories as well as the theories of elementary particle physics; it provides an explication of the vague notions of "local" and "global" gauge transformations; it explains how and why a fibre bundle structure emerges for theories which do not wear their bundle structure on their sleeves; it illuminates the connections of the gauge concept to issues of determinism and what counts as a genuine "observable"; and it calls attention to problems which arise in attempting to quantize gauge theories. Some of the limitations and problematic aspects of the formalism are also discussed.

----

1 Introduction

Given the importance of the gauge concept in modern physics, there is a surprising and non-benign neglect of this topic in the philosophical literature. There are, of course, some exceptions–e.g. the PSA 1998 symposium on field theories (see the papers by Anyang and Teller in Howard (2000)), and Margaret Morrison's recent book *Unifying Scientific Theories* (2000). But much remains to be done.

The first order of business is demystification. There is something called the "gauge argument" that goes like this. Start with a free field which admits a "global symmetry" and obeys a "global conservation law." An appeal to relativity theory and locality is then used to motivate a move from the "global" to a "local symmetry." But this move necessitates the introduction of a new field that interacts with the original field (and, perhaps, with itself) in a prescribed way. The success of the gauge argument in capturing some of the most fundamental interactions in nature has been taken to indicate that the argument reveals the logic of nature. Chris Martin's contribution to this symposium cautions against a literal reading of the gauge argument and reveals a competing logic in which the "gauge principle is the output rather than the input. will remove the veil and reveal a competing logic that makes the "gauge principle" the output rather than the input.

An even more fundamental kind of demystification is needed in getting a basic grip on the concept of gauge. Here I have two recommendations. First, don't be blinded by the glitz of the fibre bundle formalism. This formalism is arguably essential to a complete understanding of gauge. But the trouble is that this formalism is not only powerful, it is also very flexible–we can see fibre bundles all over the place. What is needed is an explanation of what the relevant fibre bundle structure is and how it arises. This need is underscored by the fact that it wasn't until quite recently that physicists began to consciously formulate theories in the fibre bundle formalism, and the majority of theories we encounter in modern physics don't wear the fibre bundle structure of their sleeves. Second, don't look only to theories that are self-consciously labeled "gauge theories." Such a focus would lead one to the wrong conclusion that gauge theories are Yang-Mills theories, whereas these theories are only one special class of gauge theories. In addition, examples drawn from more humble theories are a better way to get an initial grip on gauge.

The approach I will explore derives from the constrained Hamiltonian formalism. This formalism is widely used by one segment of the physics community but is virtually unknown to the philosophical community–this alone makes it worth airing. This approach may not be the best approach to gauge matters (if there is such a thing), but it does have a number of virtues: (i) It gives one explication of the vague talk about "local" and "global transformations"; (ii) it explains how the relevant fibre bundle structure arises (when it does); and (iii) it directly taps one of the main roots of the gauge concept and connects this concept to important foundational issues, such as the nature of observables and determinism.

Issues about observables become especially acute then one tries to quantize a theory with gauge freedom since, presumably, the quantities that get promoted to quantum observables (in the sense of self-adjoint operators on an appropriate Hilbert space) are the classical observables in the sense of gauge invariant quantities. Gordon Belot's contribution to this symposium will speak to some of the issues that arise in connection with various approaches to quantization of gauge theories. And apart from quantization, identifying the observables is a crucial element of interpreting a physical theory in the sense of saying what the world must like if the theory is to be true of the world.[1]

2 Gauge and gauge transformations: the constrained Hamiltonian formalism.

The approach I am going to describe is confined to theories whose equations of motion or field equations are derivable from an action principle.

$$\textit{Action principle}\text{: } \delta\mathcal{L} = \int_{\Omega} L(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(n)})d\mathbf{x} = \mathbf{0} \Longrightarrow \textbf{ELEs } L_A = 0,\ A = 1, 2, ..., q \tag{1}$$

(Here $\mathbf{x} = (x^1, ..., x^p)$ stands for the independent variables, $\mathbf{u} = (u^1, ..., u^q)$ stands for the dependent variables, and the $\mathbf{u}^{(n)}$ are derivatives of the dependent variables upto some finite order n with respect to the x^i.) Every such variational symmetry is a symmetry of the Euler-Lagrange equations (ELEs), i.e. carries solutions to solutions, but in general the converse is not true.

For such theories Noether's (1918) theorems provide a way to draw the distinction between local and global symmetries. Noether's first theorem concerns the case of an r-parameter Lie group $\mathcal{G}_r$, which I take to be the explication of the (badly chosen) term "global symmetry." The theorem states that the action admits a group $\mathcal{G}_r$ of variational symmetries iff there are r linearly independent combinations of the Euler-Lagrange expressions L_A, $A = 1, 2, ..., q$, which are divergences, i.e. there are r p-tuples $\mathbf{P}_j = (P_j^1, ..., P_j^p)$, $j = 1, 2, ..., r$, where the P_j^i are functions of $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{u}^{(n)}$ such that

$$Div(\mathbf{P}_j) = \sum_A c_j^A L_A, \qquad j = 1, 2, ..., r \tag{2}$$

where $Div(\mathbf{P}_j)$ stands for $\sum_{i=1}^{p} D_i P_j^i$ and D_i is the total derivative with respect to x^i. Thus, as a consequence of the ELEs, $L_A = 0$, there are r *proper conservation laws*

$$Div(\mathbf{P}_j) = 0, \qquad j = 1, 2, ..., r. \tag{3}$$

The $\mathbf{P}_j$ are called the *conserved currents*.[2] *Example*: For interacting point masses in Newtonian mechanics, the requirement that the inhomogeneous Galilean group is a variational symmetry entails the conservation of energy, angular and linear momentum, and the uniform motion of the center of mass.

Noether's second theorem is concerned with the case of an infinite dimensional group $\mathcal{G}_{\infty r}$ depending on r arbitrary functions $h_j(\mathbf{x})$, $j = 1, 2, ..., r$. This I take to be the explication of the (badly chosen) term "local symmetries." The theorem states that the action admits a group $\mathcal{G}_{\infty r}$ of variational symmetries iff there are r dependencies among the ELEs, in the form of linear combinations of derivatives of the L_A, which vanish identically. These dependencies can be interpreted as "strong" (aka "off shell") conservation laws that hold not as consequences of the ELEs but are mathematical identities (aka "generalized Bianchi identities"). Since the ELEs are not independent, we have a case of underdetermination, and as a result the solutions of these equations contain arbitrary functions of the independent variables–an apparent violation of determinism.

The underdetermination encountered in Noether's second theorem is one of the principal roots of the notions of gauge and gauge transformation. In one of its main uses, a "gauge transformation" is supposed to be a transformation that connects what are to be regarded as equivalent descriptions of the same state or history of a physical system. And one key motivation for seeking gauge freedom is to mop up the slack that would otherwise constitute a breakdown of determinism: taken at face value, a theory which admits "local" gauge symmetries is indeterministic because the initial value problem does not have a unique solution; but the apparent breakdown is to be regarded as merely apparent because the allegedly different solutions for the same initial data are to be regarded as merely different ways of describing the same evolution or, in equivalent terms, the evolution of the genuine or gauge invariant quantities is manifestly deterministic.

Now the obvious danger here is that determinism will be trivialized if whenever it is threatened we are willing to sop up the non-uniqueness in temporal evolution with what we regard as gauge freedom to describe the evolution in different ways. Is there then some non-question begging and systematic way to identify gauge freedom and to characterize the genuine observables? The answer is yes, but specifying the details involves a switch from the Lagrangian to the constrained Hamiltonian formalism. To motivate that switch, let me note that, subject to some technical provisos, if one is in the domain of Noether's second theorem (i.e. the action admits "local" symmetries = a group $\mathcal{G}_{\infty r}$ of variational symmetries), as I have been assuming is the case for gauge theories, then the Lagrangian density (more properly, its Hessian) is singular, and the Legendre transformation which defines the canonical momenta shows that these momenta are not all independent. Hence, one is in the domain of the constrained Hamiltonian theories. To illustrate what this means and to underscore the point that the key ideas about gauge arise in the humblest settings, I will work with examples drawn from classical particle mechanics.

Suppose that we are dealing with a theory whose equations of motion are derivable from the action principle

$$\delta \int L dt = 0, \qquad L = L(q, \dot{q}), \qquad \dot{q} := dq/dt. \tag{4}$$

where q notes the generalized position variables and $\dot{q}$ denotes the generalized velocity variables. (Allowance can be made for higher order Lagrangians, e.g. $L = L(q, \dot{q}, \ddot{q})$.) The familiar form the ELEs

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^n}\right) - \frac{\partial L}{\partial q^n} = 0, \quad n = 1, 2, ..., N. \tag{5}$$

can be rewritten as

$$\ddot{q}^m \left(\frac{\partial^2 L}{\partial \dot{q}^m \partial \dot{q}^n}\right) = \frac{\partial L}{\partial q^n} - \dot{q}^m \left(\frac{\partial^2 L}{\partial q^m \partial \dot{q}^n}\right). \tag{6}$$

If the Hessian matrix $\left(\frac{\partial^2 L}{\partial \dot{q}^m \partial \dot{q}^n}\right)$ is singular, one cannot solve for $\ddot{q}^m$ in terms of the positions and velocities, and determinism (apparently) fails because arbitrary functions of time appear in the solutions. But a way to recoup appears in the Hamiltonian formalism.

When the Hessian matrix $\left(\frac{\partial^2 L}{\partial \dot{q}^m \partial \dot{q}^n}\right)$ is singular, the canonical momenta defined by the Legendre transformation $p_n := \frac{\partial L}{\partial \dot{q}^n}$ are not all independent but must satisfy constraints

$$\phi_m(p, q) = 0, \qquad m = 1, 2, ..., K < N \tag{7}$$

that follow from the definitions of the momenta. These are the *primary constraints*. *Secondary constraints* may also appear when it is demanded that the primary constraints be preserved under the allowed motions. More important for our purposes is a second dichotomy which cuts across the primary vs. secondary dichotomy; namely, a constraint is *first class* if it commutes (i.e. has vanishing Poisson bracket) with all of the constraints, otherwise it is called *second class*. Dirac, who was responsible for developing this formalism, proposed that the gauge transformations be identified with the transformations generated by the first class constraints and that the observables be identified with the functions of the phase space variables (p, q) that commute with the first class constraints or, equivalently, are constant along the gauge orbits.[3]

All of this may sound scary, but a feeling for what is going can be gained from the following humble concrete illustration. Those of you who have read Maxwell's (1877) *Matter and Motion* may have been puzzled by his apparently contradictory claim that acceleration is *relative* even though rotation is *absolute*. Maxwell is consistent if we take him to be proposing that physics be done in the setting of Maxwellian spacetime. Like Newtonian spacetime Maxwellian spactime has absolute simultaneity, the $\mathbb{E}^3$ structure of the instantaneous space, and a time metric, but it eschews the full inertial structure in favor of a family of relatively non-rotating rigid frames.[4] In terms of coordinate systems adapted to the absolute simultaneity, the $\mathbb{E}^3$ structure, and the privileged non-rotating frames, the symmetry transformations of Maxwellian spacetime are

$$\begin{aligned} \mathbf{x} &\rightarrow \mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{a}(t) \\ t &\rightarrow t' + const. \end{aligned} \tag{Max}$$

where $\mathbf{R}$ is a constant rotation matrix and $\mathbf{a}(t)$ is an arbitrary smooth function of time. In such a setting it seems hopeless to have determinism if, as ordinarily assumed, positions and velocities of particles are regarded as observables. For we can choose $\mathbf{a}(t)$ such that it is zero for all $t \leq 0$ and non-zero for $t > 0$. Since a symmetry of the spacetime should be a symmetry of the equations that specify the permitted particle motions, the application of (Max) to a solution of the equations of motion will produce another solution that agrees on the particle trajectories of the first solution for all past time but disagrees with it in the future–an apparent violation of even the weakest form of determinism.

Now let's see how this example gets reinterpreted when cranked through the Lagrangian/constrained Hamiltonian formalism. An appropriate Lagrangian invariant under (Max) is

$$L = \sum \sum\nolimits_{j<k} \frac{m_j m_k}{2M} (\dot{\mathbf{r}}_j - \dot{\mathbf{r}}_k)^2 - V(|\mathbf{r}_j - \mathbf{r}_k|), \qquad M := \sum_i m_i. \tag{8}$$

The transformations (Max) are global mappings of Maxwellian spacetime onto itself, but they are "local" in that Noether's second theorem applies since (Max) contains an infinite dimensional group $\mathcal{G}_{\infty 3}$. You can verify that the Hessian matrix for (7) is singular. The ELEs are:

$$\frac{d}{dt}\left(m_i(\dot{\mathbf{r}}_i - \frac{1}{M}\sum_k m_k \dot{\mathbf{r}}_k) \right) = \frac{\partial V}{\partial \dot{\mathbf{r}}_i}. \tag{9}$$

These equations do not determine the evolution of the particle positions uniquely; for if $\mathbf{r}_i(t)$ is a solution, so is $\mathbf{r}'_i(t) = \mathbf{r}_i(t) + \mathbf{f}(t)$, for arbitrary $\mathbf{f}(t)$, confirming the intuitive argument given above for the apparent breakdown of determinism.

Now let's switch to the Hamiltonian formalism and find the constraints. The canonical momenta are:

$$\mathbf{p}_i := \frac{\partial L}{\partial \dot{\mathbf{r}}_i} = \frac{m_i}{M}\sum_k m_k(\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_k) = m_i \dot{\mathbf{r}}_i - \frac{m_i}{M}\sum_k m_k \dot{\mathbf{r}}_k. \tag{10}$$

These momenta are not independent but must satisfy three primary constraints:

$$\phi_\alpha = \sum_i p_i^\alpha = 0, \quad \alpha = 1, 2, 3. \tag{11}$$

These primary constraints are the only constraints–there are no secondary constraints. They are first class, and they generate (Max). So the genuine or gauge invariant observables of such a theory are the quantities that are invariant under (Max), which include quantities like relative positions, relative velocities, and relative accelerations of the particles. The equations of motion are deterministic with respect to these observables.

Where do fibre bundles enter this picture? In the constrained Hamiltonian formalism the *constraint surface* $\mathcal{C}$ consists of the subspace of the original phase space $\Gamma = \Gamma(p, q)$ where all of the constraints are satisfied. The reduced phase space $\tilde{\Gamma}$ is defined as the quotient of $\mathcal{C}$ by the gauge orbits generated by the first class constraints. One would like it to be the case that the gauge orbits are the fibres of a bundle with base space $\tilde{\Gamma}$. In practice this will not always be the case because various technical conditions may fail.[5] If one believes that the fibre bundle apparatus captures an essential feature of nature, then one could posit that the emergence of the appropriate bundle structure is a necessary condition for genuine physical possibility. This is an interesting idea, but obviously it requires critical examination. Turning to the opposite side of the coin, when the reduction process does go smoothly, one can wonder why the original gauge theory cannot simply the jettisoned in favor of the unconstrained Hamiltonian theory for the reduced phase space. As Gordon Belot notes in his contribution to this symposium, when the reduction process is carried out for the above toy example, one ends up with an overcomplete set of quantities (for more than three particles, relative particle distances give an over complete set of coordinates for the reduced configuration space) and writing unconstrained equations of motion requires an undemocratic choice of among these quantities. But this isn't a very strong reason to retain the gauge formalism. Perhaps a better reason for retention is to keep open the possibility of weakening or breaking one of the constraints. Again the general issue here begs for critical examination.

I want to emphasize that even for paradigm cases of gauge theories that wear their fibre bundle structure on their sleeves–e.g. Yang-Mills theories–understanding the geometry of constraints is crucial to quantization. One route to quantization is to fix a gauge and quantize in that gauge. But when one tries to do this for Yang-Mills theories using the analogues of familiar gauge conditions (e.g. Lorentz gauge) the procedure may break down. The difficulty is explained by the fact that the gauge condition fails to define a global transversal in the constraint surface, i.e. a hypersurface that meets each of the gauge orbits exactly once. Another approach to quantization is Dirac constraint quantization which promotes the first class constraints to operators and then requires that the vectors in the physical Hilbert space be annihilated by these operators. However, it can happen that the resulting quantization is inequivalent to that obtained by passing to the reduced phase space and quantizing the unconstrained Hamiltonian system! In such a case, which is the correct quantization? And how would one tell?

3 GTR and gauge

It is well-known from Einstein's notorious "hole argument" that, taken at face value, GTR is not deterministic (see Norton (1987) and Howard (1999)). The problem was redicovered in the 1950's by Peter Bergmann, who along with Dirac was responsible for developing the constrained Hamiltonian formalism:

> [I]n a general relativistic theory–and this is in sharp contrast to a Lorentz-covariant theory–no initial data enables us to predict the values of local field observables at some later time ... [Nevertheless] such a theory can be

> completely causal [i.e. deterministic], in a slightly modified sense. (1961, 510)

Namely, the theory enables us to unequivocally predict the values of *observables* from appropriate initial data. Before we can evaluate this claim we need to know what an observable in GTR is. Bergmann's first answer looked alarmingly circular: "We shall call a quantity observable if it can be predicted uniquely from initial data" (511). Fortunately there is a way out of this circle.

Einstein's field equations can be derived from an action principle. This action admits the diffeomorphism group as a variational symmetry, so one knows by Noether's second theorem that we are concerned with a case of underdetermination–another way of saying that there is an apparent violation of determinism. When one Legendre transforms to the Hamiltonian form of the theory, one discovers (no surprise) that the Hamiltonian form of Einstein's field equations constitutes a constrained Hamiltonian system. Thus, one can proceed as above define an *observable* for GTR to be a dynamical quantity that is gauge invariant in Dirac's sense, i.e. has vanishing Poisson bracket with all of the first class constraints or, equivalently, is constant along the gauge orbits.

But there are two surprises awaiting us. The first is that the Lie algebra of the constraints in GTR is not closed; that is to say, it is not a genuine Lie algebra. Since the closure is a defining feature of Yang-Mills theories, it follows that GTR is not a Yang-Mills theory. Now some writers want to reserve the label "gauge theory" for Yang-Mills theories. This seems to me to be a merely terminological matter–if you do not wish to call GTR a gauge theory because it is not Yang-Mills, that is fine with me; but please be aware that the constrained Hamiltonian formalism provides a perfectly respectable sense in which the standard textbook formulation of GTR using tensor fields on differentiable manifolds does contain gauge freedom. What goes beyond label mongering is the issue of why GTR fails to be a Yang-Mills theory and, more generally, what features separate constrained Hamiltonian theories that are Yang-Mills from those which are not. Some important results of these matters have been obtained by Lee and Wald (1990), but these results are too technical to review here.

The second and even more unpleasant surprise is that taking the Bergmann route leads to a revival of McTaggartism. There are two families of constraints in the constrained Hamiltonian formulation of GTR: the *momentum constraints* and the *Hamiltonian constraints*. The former generates diffeomorphisms in the initial value surface while the latter generates motion. Since motion is pure gauge, it follows that all of the gauge invariant quantities of the theory are constants of the motion. Insofar as genuine physical quantities are gauge invariant, they do not change with time. Reactions in the physics literature to this result are divided. One side there are those (e.g. Kuchař (1993)) who say that the result is incorrect: obviously (they claim) there is genuine change, which means that the Hamiltonian constraints in GTR cannot be taken to generate gauge. Those who choose this route have to explain how they propose to implement general covariance and how they escape the indeterminism of the hole argument. On the other side there are those (e.g. Rovelli (1991)) who accept the result and try to live with it. Those who choose the second route have to explain how to do physics–in particular, how to quantize and how to do statistical mechanics without unfreezing the dynamics and introducing a time variable with an associated non-zero Hamiltonian. They also owe us an explanation of how the illusion of change arises in an unchanging world. I won't prejudge the issue and will only say that here we have an example of how following out the recommended approach to gauge leads to some fundamental problems in the foundations of physics that are of considerable philosophical interest.[6]

4 Problems and prospects

I have touted the constrained Hamiltonian formalism as a kind of royal road to the understanding of gauge. It is now time to take notice of some potholes and potential dead ends in this road.

The scope of the approach is very broad in the sense that it applies to the vast majority of theories in modern physics. But this is because physicists, for one reason or another, have concentrated on theories where the field equations or equations of motion are derivable from an action principle. There are notable exceptions, such as the Cartan formulation of Newtonian gravitation, where one would like be able to speak of gauge (see Norton (1995)).

A more serious challenge is concerned not with the boundaries of the domain of applicability of the constrained Hamiltonian formalism but rather with the fact that within its domain of application it can render different verdicts on what counts as an observable depending on the way in which the theory is formulated. A relevant example comes from Janis (1969). In the conventional formulation of Maxwell's theory of electromagnetism in Minkowski spacetime the electromagnetic field tensor F_{ab} is written in terms of a four-potential A_a, $a = 1, 2, 3, 4$:

$$F_{ab} := 2A_{[a,b]} \tag{12}$$

where the square brackets denote anti-symmetrization. Maxwell's equations in vacuo are

$$F^{ab}{}_{,b} = 0 \tag{13a}$$
$$^{*}F^{ab}{}_{,b} = 0 \tag{13b}$$

where the $*$ denotes the dual tensor and indices have been raised using the Minkowski metric. The theory can be recast in terms of two four-potentials, A_a and $\bar{A}_a$ by writing

$$F_{ab} = 2(A_{[a,b]} +^* \bar{A}_{[a,b]}). \tag{14}$$

With the Lagrangian density $L := \frac{1}{2}F_{ab}F^{ab}$ per usual, this new version of Maxwell's theory also yields the field equations (13). In the first version, the field tensor, but not the potentials, are observables. But in the second version, $\bar{A}^{[a,b]}$ as well as ${}^*A^{[a,b]}$ is an observable.

My response is to see different theories rather than different formulations of the same theory. A theory is a set of equations *plus* an intended interpretation, part of which involves a specification of what is to count as an observable. The constrained Hamiltonian formalism constrains this specification. But no formalism–the constrained Hamiltonian formalism or other–can tell you which theory is "best." That has to be decided by a combination of experimental and theoretical considerations. At the present time there is no reason to think that ${}^*\bar{A}^{[a,b]}$ is needed to characterize the electromagnetic field. But if magnetic monopoles were detected, the situation would be different. (Electric charges and currents would serve as the sources for $A^{[a,b]}$ while magnetic charges and currents would serve as sources for ${}^*\bar{A}^{[a,b]}$.)

Let me now turn to a worry about the status of determinism. I said that determinism becomes a trivial doctrine if whenever cracks appear in the doctrine we are ready to paper over the cracks by seeking gauge freedom. And then I gave the impression that the trivialization is halted by providing a principled way to detect gauge freedom. This impression is badly misleading if the means of detecting gauge freedom is that of Dirac. Start with any theory derivable from an action principle, and suppose that the ELEs do not suffer from overdetermination but do suffer from underdetermination–they fail to determine a unique solution from initial data because arbitrary functions of time appear in the solutions. A cure for this form of indeterminism is always at hand in that in the constrained Hamiltonian formalism the gauge transformations as identified by Dirac's prescription are sufficient to sop up the underdetermination. (Of course, it may happen that the cure takes the drastic form of freezing the dynamics!) To be sure, the assumption that the field equations or equations of motion be derivable from an action principle is a strong restriction, but nevertheless I find it disconcerting that for such cases a cure for underdetermination is always possible.

Finally, I need to say something about Dirac's proposal (or conjecture as it is sometimes called) that *all* first class constraints, secondary, tertiary, etc. generate gauge transformations. The alleged counterexamples are rather technical and artificial. But this issue cannot be dismissed as a mere technical problem since it gets at the heart of what counts as a gauge transformation. The counterexamples concern cases where Dirac's conjecture would count some degree of freedom as pure gauge even though the ELEs determine that degree of freedom. So the need to mop up underdetermination is not present as a motivation for seeing gauge freedom. This raises the issue of whether gauge freedom and observables be characterized directly in terms of the Lagrangian formalism without having to resort to the constrained Hamiltonian formalism. As an example of what I have in mind, suppose that the action $\mathcal{L} = \int L(q, \dot{q})dt$ admits the group $\mathcal{G}_{\infty r}$ of variational symmetries. Define the observables to be functions of the velocity phase space $(q, \dot{q})$ which are invariant under $\mathcal{G}_{\infty r}$. Is this notion of observable equivalent, in some relevant sense, to Dirac's notion? If not, which concept of observable is preferable?

5 Conclusion

The constrained Hamiltonian formalism provides one window on the gauge concept. I don't say that we should look only through this window; indeed, let a thousand windows be opened. But I do say that the approach is sufficiently promising that it deserves more attention from philosophers of science; in particular, exploring the gauge concept by means of this formalism reveals a number of fundamental interpretational problems concerning the status of determinism, the nature of observables, McTaggartism, etc. The gauge concept is important not only for understanding Yang-Mills theories and theories of elementary particle physics but also for Newtonian and classical relativistic theories as well. We will need a big toolbox to build an analysis that is adequate to all of the various senses and applications of the gauge concept across the broad range of physical theories. Fibre bundles will surely be in the toolbox, but much else besides including, I believe, the constrained Hamiltonian formalism.[7]

Notes

1. 'Observable' is being used as a term of art to denote a fundamental physical magnitude. The values of such a quantity may or may not be directly observable by the unaided senses. What conditions a quantity must satisfy in order to count as an observable for a theory T is an unresolved problem. But for gauge theories a necessary condition is that the quantity is gauge invariant.

2. In some cases this conservation law can be written in the form $D_t T + div(X) = 0$, where *div* is the spatial divergence. Then if the flux density X vanishes on the spatial boundary of the system, the spatial integral of the density T is a constant of the motion.

3. A detailed treatment of the constrained Hamiltonian formalism is to be found in Henneaux and Teitelboim (1992).

4. See my (1989) for details.

5. Suppose that the first class constraints form a Lie algebra and that this algebra exponentiates to give a Lie group $\mathcal{G}$ which acts freely on $\mathcal{C}$. And suppose that the quotient $\mathcal{C}/\mathcal{G}$ is a manifold. Then $\mathcal{C}/\mathcal{G}$ will be the base space of a $\mathcal{G}$-bundle whose fibres are the orbits of the group action. Not all of these conditions are necessary for the gauge orbits generated by the first class constraints to be the fibres of a nice bundle. I don't know what a minimally necessary condition is.

6. Belot and Earman (2000) defends the strategy of swallowing modern McTaggartism and exploring the implications for physics and philosophy of time.

7. A more complete version of this paper can be found on the PSA Y2K website.

References

Belot, G and Earman, J. 2000. "Thoroughly Modern McTaggart: Or What if McTaggart Had Learned Einstein's General Theory of Relativity," pre-print.

Bergmann, P. G. 1961. "Observables in General Relativity," *Reviews of Modern Physics* **33**: 510-514.

Earman, J. 1989. *World Enough and Spacetime: Absolute vs. Relational Theories of Space and Time*. Cambridge, MA: MIT Press.

Henneaux, M. and Teitelboim, C. 1992. *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press.

Howard, D. 1999. "Point Coincidences and Pointer Coincidences: Einstein and the Invariant Content of Space-Time Theories." In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7), 463-500. Boston: Birkhäuser.

Howard, D. A. 2000. *PSA 98*. Supplement to Volume 67, Number 3 of *Philosophy of Science*.

Janis, A. I. 1969. "Energy and Observables in Electromagnetic theory," *American Journal of Physics* **38**: 202-206.

Kuchař. K. 1993. "Canonical Quantum Gravity," in R. J. Gleiser, C. N. Kozameh, and O. M. Moreschi (eds.), *General Relativity and Gravitation 1992*, 119-150. Philadelphia, PA: IOP Publishing.

Lee, J. and Wald, R. M. 1990. "Local Symmetries and Constraints," *Journal of Mathematical Physics* **31**: 725-743.

Maxwell, J. C. 1877. *Matter and Motion*. Reprinted by Dover Books (New York, 1951).

Morrison, M. 2000. *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

Noether, E. 1918. "Invariante Variationsprobleme," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse.* Heft 2: 235-257.

Norton, J. D. 1987. "Einstein, the Hole Argument and the Reality of Space." In J. Forge (ed.), *Measurement, Realism and Objectivity*, 153-188. Boston: D. Reidel.

Norton, J. D. 1995. "The Force of Newtonian Cosmology: Acceleration is Relative," *Philosophy of Science* **62**: 511-522.

Rovelli, C. 1991. "Time in Quantum Gravity: An Hypothesis," *Physical Review D* **43**: 442-456.