8

Tracking down gauge: an ode to the constrained Hamiltonian formalism

JOHN EARMAN

1 Introduction

Like moths attracted to a bright light, philosophers are drawn to glitz. So in discussing the notions of 'gauge', 'gauge freedom', and 'gauge theories', they have tended to focus on examples such as Yang–Mills theories and on the mathematical apparatus of fibre bundles. But while Yang–Mills theories are crucial to modern elementary particle physics, they are only a special case of a much broader class of gauge theories. And while the fibre bundle apparatus turned out, in retrospect, to be the right formalism to illuminate the structure of Yang–Mills theories, the strength of this apparatus is also its weakness: the fibre bundle formalism is very flexible and general, and, as such, fibre bundles can be seen lurking under, over, and around every bush. What is needed is an explanation of what the relevant bundle structure is and how it arises, especially for theories that are not initially formulated in fibre bundle language.

Here I will describe an approach that grows out of the conviction that, at least for theories that can be written in Lagrangian/Hamiltonian form, gauge freedom arises precisely when there are Lagrangian/Hamiltonian constraints of an appropriate character. This conviction is shared, if only tacitly, by that segment of the physics community that works on constrained Hamiltonian systems.[1] The approach taps into one of the root notions of gauge transformations – namely, that of transformations that connect equivalent descriptions of the same state or history of a physical system – and one of the key motivations for seeking gauge freedom – namely, to take up the slack that would otherwise constitute a failure of determinism.

[1] Here is one explicit expression of that conviction: 'It is well known that all the theories containing gauge transformations are described by constrained systems' (Gomis, Henneaux, and Pons, 1990, p. 1089). The conviction is all but explicit in Henneaux and Teitelboim (1992), the standard reference on the quantization of gauge systems, which opens with a long and detailed treatment of constrained Hamiltonian systems. In section 7 below I will indicate a way in which this conviction can be challenged.

2 Noether's theorems, constrained Hamiltonian systems, and all that

The literature on gauge theories is filled with talk about 'global' and 'local' symmetries, talk which is annoying both because it is often unaccompanied by any attempt to make it precise and because it is potentially very misleading (e.g. a global mapping of the spacetime onto itself can count as a 'local' symmetry in the relevant sense of corresponding to a gauge transformation). One way to get a grip on the global vs. local distinction is to place it in the context of Noether's two theorems.[2] Both theorems apply to theories whose equations of motion or field equations are derivable from an action principle and, thus, are in the form of (generalized) Euler–Lagrange (EL) equations. And both concern *variational symmetries*, that is, a group $\mathcal{G}$ of transformations that leave the action $\mathcal{A} = \int_\Omega L(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(n)})\, d^p\mathbf{x}$ invariant, where $\mathbf{x} = (x^1, \ldots, x^p)$ stands for the independent variables, $\mathbf{u} = (u^1, \ldots, u^r)$ are the dependent variables, and the $\mathbf{u}^{(n)}$ are derivatives of the dependent variables up to some finite order n with respect to the x^i.[3] Every such variational symmetry is a symmetry of the EL equations $L_A = 0$, $A = 1, 2, \ldots, r$; that is, a variational symmetry carries solutions of the EL equations into solutions. The converse, however, is not guaranteed to hold, e.g. it is often the case that scaling transformations are symmetries of the EL equations but are not variational symmetries.

Noether's first theorem concerns the case of an s-parameter Lie group $\mathcal{G}_s$, which I take to be the explication of the (badly chosen) term 'global symmetry'. The theorem states that the action admits such a group $\mathcal{G}_s$ of variational symmetries iff there are s linearly independent combinations of the EL expressions L_A which are divergences, i.e. there are s p-tuples $\mathbf{P}_j = (P_j^1, \ldots, P_j^p)$, $j = 1, 2, \ldots, s$, where the P_j^i are functions of $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{u}^{(n)}$ such that

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

where the c_j^A are constants, $Div\,(\mathbf{P}_j)$ 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 EL equations there are s *conservation laws*

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

[2] These theorems were presented in Noether (1918). For relevant historical information, see Kastrup (1987) and Byers (1999). For a modern presentation of the Noether theorems, see Olver (1993), Brading (2002), and Brading and Brown (this volume).

[3] The transformations $\mathcal{G} \ni g : (\mathbf{x}, \mathbf{u}) \to (\mathbf{x}', \mathbf{u}')$ may depend on both the independent variables $\mathbf{x}$ and the dependent variables $\mathbf{u}$. Noether's theorems can be generalized to handle transformations that depend on the $\mathbf{u}^{(n)}$ as well (see Olver, 1993), but these generalized transformations will play no role here. But what is relevant here is the fact that Noether's original theorems can be generalized to handle so-called divergence (variational) symmetries that leave the action invariant only up to a term of the form $\int_{\partial\Omega} Div\,(\mathbf{B})$, where the variation of $\mathbf{B}$ vanishes on the boundary $\partial\Omega$. For example, the Galilean velocity boosts do not leave the familiar action for Newtonian particle mechanics invariant, but these boosts are divergence (variational) symmetries. This is crucial in deriving the constancy of the velocity of the centre of mass of the system. Noether does not seem to have been aware of this fact. From here on when I speak of variational symmetries I will mean divergence (variational) symmetries.

The $\mathbf{P}_j$ are called the *conserved currents*. In some cases these conservation laws 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. It is well to note that there are equations of motion (or field equations) which cannot be derived from an action principle, and in such cases there is no guarantee that a symmetry of the equations of motion will give rise to a corresponding conserved quantity.[4]

A concrete application of Noether's first theorem is provided by interacting point masses in Newtonian mechanics, provided that the equations of motion follow from an action principle. 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 centre of mass. Conversely, the existence of these conservation laws entails that the action admits a 10-parameter Lie group of variational symmetries.

Noether's second theorem is concerned with the case of an infinite-dimensional Lie group $\mathcal{G}_{\infty s}$ depending on s arbitrary functions $h_j(\mathbf{x})$, $j = 1, 2, \ldots, s$, of all of the independent variables. This I take to be the explication of the (badly chosen) term 'local symmetries'. The theorem states that the action admits such a group $\mathcal{G}_{\infty s}$ of variational symmetries iff there are s dependencies among the EL equations, in the form of linear combinations of the L_A and their derivatives, which vanish identically (also known as 'generalized Bianchi identities'). Since the EL equations 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 since the initial data do not seem to fix a unique solution of the EL equations.

Now comes an insight of Noether's that is important enough to be labelled the third Noether theorem. Suppose that $\mathcal{G}_{\infty s}$ possesses a rigid Lie subgroup $\mathcal{G}_s$ that arises from fixing $h_j(\mathbf{x}) = \alpha_j = const$. Then each of the conserved currents $\mathbf{P}_j$, corresponding to the invariance under $\mathcal{G}_s$ by Noether's first theorem, can be written as a linear combination of the EL expressions L_A plus the divergence of an antisymmetric quantity: $P_j^i = \sum_{A=1}^{r} L_A \xi_j^{Ai} + \sum_{k=1}^{p} D_k X_j^{ik}$, where the 'superpotentials' X_j^{ik} are functions of $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{u}^{(n)}$, and $X_j^{ik} = -X_j^{ki}$. Thus, on any solution to the EL equations $Div\,(\mathbf{P}_j) = \sum_{i,k=1}^{p} D_i D_k X_j^{ik} \equiv 0$. In this case the conservation laws (2) were dubbed 'improper' by Noether. If the independent variables are spacetime coordinates and x^4 is the time coordinate, the charges Q_j associated with the conserved currents of improper conservation laws are defined by the spatial volume

[4] See section 7 below for a discussion of the issue of what equations of motion can be derived from an action principle.

integrals $\int_{V_3} P_j^4 \, d^3x = \int_{V_3} \sum_{k=1}^{4} D_k X_j^{4k} \, d^3x = \int_{V_3} \sum_{k=1}^{3} D_k X_j^{4k} \, d^3x$, which are equal to $\int_\sigma \sum_{k=1}^{3} n_k X_j^{4k} \, d\sigma$, where σ is the 2-surface bounding the 3-volume V_3, and n_k is the unit normal to the surface. In electromagnetism this relation expresses the charge in a spatial volume as the flux of the electric field through the bounding surface. The upshot is that the content of improper conservation laws amounts to a Gauss-type relation.[5] The conservation law associated with the conserved current $\mathbf{P}_j$ will be completely trivial if $\mathbf{P}_j$ vanishes on any solution of the EL equations; this happens just in case the total divergence of the superpotential for $\mathbf{P}_j$ vanishes. Finally, it is worth mentioning that there is a converse for Noether's third theorem; namely, if corresponding to a non-trivial variational symmetry there is a conserved quantity that can be written as a linear combination of the EL expressions and the divergence of a superpotential, then there is an infinite dimensional Lie group of variational symmetries and the EL equations are underdetermined (see Olver, 1993, section 5.3).

In cases where Noether's second theorem applies there are also what physicists call 'strong conservation laws' which hold 'off shell', i.e. regardless of whether the EL equations hold or not. These strong laws, which are consequences of the Bianchi identities, should not be confused with improper conservation laws.

To return to the main theme: the underdetermination encountered in Noether's second theorem points to 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 take 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. Putting the point in different terminology, the evolution of the genuine or gauge-invariant quantities (or 'observables') is manifestly deterministic.

Now the obvious danger here is that determinism will be trivialized if, whenever it is threatened by non-uniqueness, we stand 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

[5] For an interesting discussion of how the distinction between proper and improper conservation laws illuminates Hermann Weyl's work on gauge theories, see Brading (2002).

to identify gauge freedom and to characterize the 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 $\mathcal{G}_{\infty s}$ of variational symmetries), as I have been assuming is the case for gauge theories, then the Lagrangian density (more properly, its Hessian) is singular (see Wipf, 1994), and the Legendre transformation which defines the canonical momenta shows that these momenta are not all independent but must satisfy a family of constraints. Hence, one is in the domain of the constrained Hamiltonian theories. Following Dirac (1950; 1951; 1964) one then identifies gauge transformations as mappings of the phase space that are generated by a subset of the constraints.[6] To illustrate what this means and to underscore the point that the key ideas about gauge arise in the humblest settings, I will concentrate in the following section on finite-dimensional systems.

3 Gauge transformations and first-class Hamiltonian constraints[7]

In this section I confine attention to systems where $\dim(Q) = N < \infty$, Q being the configuration space. I further restrict attention to first-order Lagrangians $L(q^i, \dot{q}^i)$, $i = 1, 2, \ldots, N$, $\dot{q}^i := \frac{dq^i}{dt}$, and assume in addition that the Lagrangians are independent of the time t. In this setting the EL equations assume their familiar form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^n}\right) - \frac{\partial L}{\partial q^n} = 0, \qquad n = 1, 2, \ldots, N \tag{3}$$

These equations can be rewritten as

$$\ddot{q}^m \left(\frac{\partial^2 L}{\partial \dot{q}^m \partial \dot{q}^n}\right) = \frac{\partial L}{\partial q^n} - \dot{q}^m \left(\frac{\partial^2 L}{\partial q^m \partial \dot{q}^n}\right). \tag{4}$$

When the Hessian matrix $W_{ij} := (\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j})$ 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 and, thus, initial data will not single out a unique solution. A way to recoup the fortunes of determinism appears in the Hamiltonian treatment. But before turning to that treatment, I need to define another structure.

Setting $v^k := \dot{q}^k$ and $\alpha_i := \frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial v^i \partial q^k} v^k$, the EL equations (4) can be rewritten as

$$\ddot{q}^m W_{mn} = \alpha_i. \tag{5}$$

[6] The constrained Hamiltonian formalism was developed simultaneously by Peter Bergmann and co-workers; see, for example, Bergmann (1949) and Bergmann and Brunings (1949).

[7] The general reference for this section is Henneaux and Teitelboim (1992).

So if the null vector fields of the Hessian are $V_\rho := \gamma^i_\rho \frac{\partial}{\partial v^i}$, the first-generation Lagrangian constraints are

$$\chi^{(1)}_\rho := \alpha_i \gamma^i_\rho = 0, \qquad \rho = 1, 2, \ldots, R < N. \tag{6}$$

Requiring that these constraints be preserved by the motion produces a second generation of Lagrangian constraints, etc. Eventually this process terminates. The final Lagrangian constraint manifold $\mathcal{C}^\ell_F$ is then the submanifold of the Lagrangian velocity phase space $\Gamma(q, v) := T(Q)$ where all of the constraints hold.

Passing to the Hamiltonian phase space $\Gamma(q, p) := T^*(Q)$ is accomplished by the Legendre transformation $FL : \Gamma(q, v) \to \Gamma(q, p)$, where $FL(q^i, v^i) = (q^i, p_i)$ with the canonical momenta given by $p_i := \frac{\partial L}{\partial v^i}$. When the Hessian is singular, the canonical momenta are not all independent but must satisfy *primary constraints*

$$\phi^{(0)}_\mu(q, p) = 0, \qquad \mu = 1, 2, \ldots, J < N \tag{7}$$

that follow from the definitions of the momenta. These equations define the primary Hamiltonian constraint manifold $\mathcal{C}^h_o$. The Hamilton–Dirac (HD) equations of motion take the form

$$\dot{q}^i \underset{\mathcal{C}^h_o}{=} \{q^i, H_T\} \tag{8a}$$

$$\dot{p}_i \underset{\mathcal{C}^h_o}{=} \{p_i, H_T\}. \tag{8b}$$

Here $\underset{\mathcal{C}^h_o}{=}$ means equality on $\mathcal{C}^h_o$ and $\{,\}$ is the usual Poisson bracket. The total Hamiltonian H_T is $H_c(q, p) + \lambda^\mu \phi^{(0)}_\mu(q, p)$, where the canonical Hamiltonian H_c is any function of (q, p) satisfying $H_c(q, p) = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L$. The Lagrange multipliers λ^μ appearing in the total Hamiltonian can be arbitrary functions of time. Requiring that the primary constraints be preserved by the motion can lead to *secondary constraints*. Requiring that the secondary constraints be preserved by the motion can produce *tertiary constraints*, etc. The final Hamiltonian constraint manifold $\mathcal{C}^h_F$, which is reached after a finite number of steps, is the submanifold of $T^*(Q)$ where all the constraints are satisfied. A constraint is said to be (*final*) *first class* just in case its Poisson bracket with any constraint 'vanishes weakly', i.e. is zero when evaluated on $\mathcal{C}^h_F$. On $\mathcal{C}^h_F$ where (8a) and (8b) have solutions, the total Hamiltonian can be written as $H_T = H^F_c + u^{\mu_f}\psi^{(0)}_{\mu_f}$, where $\psi^{(0)}_{\mu_f}$, $\mu_f = 1, 2, \ldots, K$, are the final primary first-class constraints and H^F_c is a particular canonical Hamiltonian which happens to be a first-class function, i.e. its Poisson bracket with every constraint vanishes weakly.

Under the regularity conditions that the Legendre map has constant rank and that the rank of the Poisson bracket of constraints is constant, the Hamiltonian and Lagrangian treatments of constrained systems are equivalent in the sense that for any solution $q^i(t)$ of the EL equations, there is a solution $(q^i(t), p_i(t))$ of the HD equations, and vice versa (see Batlle *et al.*, 1986).

We are now in a position to be more specific about what counts as a gauge transformation in the Hamiltonian formalism. The first version treats a gauge transformation as a point transformation on the Hamiltonian phase space $\Gamma(q, p)$.

Def. 1. (Hamiltonian version). $(q_{1,}\ p_1), (q_{2,}\ p_2) \in \mathcal{C}_F^h$ are Hamiltonian gauge equivalent iff there is a $(q_{0,}\ p_0) \in \mathcal{C}_F^h$ such that $(q_{1,}\ p_1)$ and $(q_{2,}\ p_2)$ are both obtained from $(q_{0,} p_0)$ as solutions to the HD equations in the same lapse time Δt.

The motivation for this definition starts with the conviction that the physical state of the system at any given time is completely specified by a point in phase space, and then proceeds to the realization that, if the EL/HD equations for a constrained system are to determine the physical state at later times, the physical-state-to-phase-space correspondence must be one–many.[8] The slack of the 'many' side is identified as gauge freedom. Using Def. 1 we can proceed to relate the gauge freedom to the constraints. Note first that any (final) first-class constraint can be obtained from an iterative procedure that starts with the primary first-class constraints and successively takes the Poisson bracket of the preceding stage with H_c^F. It is natural to conjecture that (i) all (final) first-class constraints generate Hamiltonian (point) gauge transformations, and (ii) any (point) Hamiltonian gauge transformation is generated by the (final) first-class constraints. This will be true provided that the ancestry of all the first-class constraints is untainted in the sense that the iterative procedure which generates the first-class constraints does not pass through an *ineffective constraint*, i.e. a constraint whose gradient vanishes weakly (see Cabo and Louis-Martinez, 1990).[9] Geometrically the gauge orbits on $\mathcal{C}_F^h$ are the integral curves of the vector field formed by linear combinations of the gradients of the first-class constraints. This vector field is null with respect to the pre-symplectic form induced on $\mathcal{C}_F^h$ by the symplectic form for $\Gamma(q, p)$.[10]

With the proviso that ineffective constraints are absent, the results (i) and (ii) justify the Dirac slogan 'The gauge transformations are those transformations

[8] As Henneaux and Teitelboim put it: '[A]lthough the physical state is uniquely defined once the set of p's and q's is given, the converse is not true – i.e., there is more than one set of values of the canonical variables representing a given physical state. To see how this conclusion comes about, we notice that if we give an initial set of canonical variables at time t_1 and thereby completely define the physical state at that time, we expect the equations to *fully determine the physical state at other times*. Thus, by definition any ambiguity in the value of the canonical variables at $t_2 \neq t_1$ should be a physically irrelevant ambiguity' (1992, pp. 16–17).

[9] For examples of what can go wrong when ineffective constraints are present, see Henneaux and Teitelboim (1992, pp. 19–20) and Gotay and Nester (1984).

[10] The mathematically precise way to formulate Hamiltonian mechanics uses a *symplectic form* ω that is a non-degenerate two-form on the phase space $T^*(Q)$. The Poisson bracket for phase functions is defined by $\{f, g\} := \omega(df, dg)$. Locally, coordinates (q^i, p_i), $i = 1, 2, \ldots, N$, can be chosen so that $\{f, g\} = \sum_i \left(\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} \right)$.

generated by the first-class constraints.' Adding non-primary first-class constraints with their Lagrange multipliers to the total Hamiltonian results in what is called the *extended Hamiltonian*. This extension is harmless (and pointless) when ineffective constraints are absent. But if ineffective constraints are present the extension can be pernicious in that it can break the equivalence of the Lagrangian and Hamiltonian treatments.[11]

There is, of course, a Lagrangian version of Def. 1. The equivalence of the Lagrangian and Hamiltonian approaches means that the Lagrangian and Hamiltonian (point) gauge transformations have a natural correspondence. So if there are no ineffective Hamiltonian constraints, every first-class Hamiltonian constraint generates a Hamiltonian (point) gauge transformation which has a corresponding Lagrangian (point) gauge transformation, and every Lagrangian (point) gauge transformation has a corresponding Hamiltonian (point) gauge transformation generated by a first-class constraint.

Given these nice equivalences of the Lagrangian and Hamiltonian treatments, it is not surprising that, despite the fact that the final Lagrangian and Hamiltonian constraint manifolds can have different dimensions, the two treatments give the same counts for the number of physical degrees of freedom. Count these degrees of freedom in the Hamiltonian and Lagrangian approaches respectively by the dimensions of $\mathcal{C}_G^h \subset \mathcal{C}_F^h$ and $\mathcal{C}_G^\ell \subset \mathcal{C}_F^\ell$, where the subscript G indicates the submanifold obtained by killing the gauge freedom by gauge fixing conditions. If there are no ineffective constraints, $\dim(\mathcal{C}_G^h) = \dim(\mathcal{C}_G^\ell) = 2N - M - P$, where N is the dimension of the configuration space, M is the number of Hamiltonian constraints, and P is the number of first-class Hamiltonian constraints (see Gràcia and Pons, 1988).

It remains to link the notion of gauge transformation detailed above with the Noether transformations. There appears to be a mismatch since the former are point transformations while the latter are mappings of solutions onto solutions. To bridge the gap, one can introduce a notion of gauge transformation connecting solutions. The Hamiltonian version is given in:

Def. 2. Two solutions $(q(t), p(t))$ and $(q'(t), p'(t))$ of the HD equations are gauge equivalent just in case for each t the points of phase space they determine are gauge equivalent in the sense of Def. 1.

A mapping $G : T(Q)^* \times \mathbb{R} \to \mathbb{R}$ of the form $G(q, p, t) = \sum_{k\geq 1} \varepsilon^k(t) G_k(q, p)$,

[11] Thus, without the restriction of no ineffective constraints, 'Dirac's conjecture' (as it is called in the literature) that all first-class constraints generate gauge transformations is false.

where $\varepsilon^k(t)$ is the kth time derivative of the gauge generator $\varepsilon(t)$, is a solution gauge transformation provided that the G_k satisfy the chain condition

$$\begin{aligned} G_1 &\equiv PFC \\ \{PFC, G_k\} &\equiv PFC \\ \{G_k, H\} + G_{k+1} &\equiv PFC \end{aligned} \tag{9}$$

where 'PFC' stands for a primary first-class constraint and H is a first-class Hamiltonian, and provided that the maximum value of k is the number of steps in the Dirac consistency algorithm (see Gràcia and Pons, 1988; Gomis *et al*., 1990).[12]

Corresponding to this transformation there will be a Lagrangian gauge transformation that maps any solution to the EL equations to another solution that is gauge equivalent to the first in the sense analogous to Def. 2. However, there is no guarantee that this Lagrangian gauge transformation is a Noether transformation in the sense that it leaves the action invariant (up to a total divergence). But it is known that the latter will hold if all of the Hamiltonian constraints are primary first class. In the opposite direction, a pair of gauge equivalent solutions to the EL equations has a corresponding pair of gauge equivalent solutions to the HD equations where the solution gauge transformation connecting them is of the form given above.

The final concept I will introduce for the Hamiltonian formalism is that of an *observable* or gauge independent quantity. There are two equivalent ways to make this precise: an observable can be defined to be a function from the Hamiltonian phase space $\Gamma(q, p)$ to $\mathbb{R}$ which has weakly vanishing Poisson brackets with the first-class constraints or, equivalently, which is constant along the gauge orbits. An observable in this sense corresponds to a function on the reduced phase space obtained by quotienting $\Gamma(q, p)$ by the gauge orbits.

The reader who is encountering the apparatus of constrained systems for the first time may be aghast at the seeming complexity. I have no comfort to offer. For not only have I slurred over some of the complexities in my overly brief presentation above, but I have additional bad news as well. Many of the systems one wants to study in physics have infinite-dimensional configuration spaces, and the relatively simple results reported above for finite-dimensional systems do not always carry over to the infinite-dimensional case. Nevertheless, in what follows I will simply forge ahead under the assumption that in constrained systems gauge is generated by the first-class Hamiltonian constraints.[13]

In closing this section I want to underscore the point that the use of the recommended apparatus for getting a fix on gauge takes the overcoming of an apparent

[12] The restriction to the case of no ineffective constraints is also in force here.

[13] As far as I am aware there are no proofs in the literature of the infinite-dimensional analogues of all of the basic results quoted above for finite dimensional systems. Belot (2002) provides a detailed discussion of a large class of gauge theories that arises when the conserved quantities of a Hamiltonian system are set equal to zero. This class contains Yang–Mills type theories but not all constrained Hamiltonian systems.

underdetermination to be the key motivation for recognizing gauge freedom. It is well to note, however, that other motivations can operate, and these motivations may produce a different verdict on gauge. Suppose, for example, that you want to be a relationist about space and time, and also that you want to acknowledge the striking success in the use of Minkowski spacetime for formulating the theories of modern physics. You could reconcile these two desires by saying that the relational spatiotemporal structure of physical events conforms to those prescribed by Minkowski spacetime while at the same time denying that physical events are in any literal sense located in a spacetime container. But to make such a stance consistent requires treating a Poincaré boost of the matter fields on Minkowski spacetime as a gauge transformation in the sense that it produces not a different physical situation but a different representation of the same physical situation. By contrast, an application of the constraint apparatus to Maxwell electromagnetic theory and other standard special relativistic theories does not produce the verdict that the Poincaré group is a gauge group. The difference between the two motivations for seeing gauge freedom is brought out by the contrasting attitudes towards what counts as an observable. A non-relationist who is guided by the constrained Hamiltonian formalism sees no interesting gauge freedom in the familiar special relativistic treatment of electromagnetism and, therefore, will treat the electromagnetic field as an observable. By contrast, the relationist is (on my reading) committed to denying that the electromagnetic field is a genuine physical observable since, by her lights, this field is not gauge invariant; and by her lights, only non-local quantities, such as the spacetime volume integral of the field energy, will pass muster. Lest one think that this result is a *reductio* of relationism, the relationist can note that worse is to follow in the context of the General Theory of Relativity (GTR). But in contrast to special relativistic theories, there is no choice to be made in GTR where the constraint structure of the theory dictates that local fields are not observables (see section 5).

4 A toy example

Since the discussion thus far has been both abstract and complicated, it may assist the reader to work through a concrete toy example that illustrates some of the above concepts. 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* (see sections 32–35 and 104–105). Maxwell is consistent if we take him to be proposing that physics be done in the setting of what I have dubbed *Maxwellian spacetime*. Like Newtonian spacetime, Maxwellian spacetime has absolute simultaneity, the $\mathbb{E}^3$ structure of the instantaneous space, and a time metric, but it eschews the full inertial structure in favour of a family of relatively non-rotating rigid frames.[14] 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

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

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

$$L = \sum\sum\nolimits_{j<k} \frac{m_j m_k}{2M}(\dot{\mathbf{r}}_j - \dot{\mathbf{r}}_k)^2 - V(|\mathbf{r}_j - \mathbf{r}_k|), \qquad M := \sum_i m_i. \tag{10}$$

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 $\mathcal{G}_{\infty 3}$ whose parameters are arbitrary functions of t, the only independent variable in the action $\int L\, dt$. It is easy to verify that the Hessian matrix for (10) is singular. The EL equations are:

$$\frac{d}{dt}\left(m_i\left(\dot{\mathbf{r}}_i - \frac{1}{M}\sum_k m_k \dot{\mathbf{r}}_k\right)\right) = \frac{\partial V}{\partial \dot{\mathbf{r}}_i}. \tag{11}$$

These equations do not determine the evolution of the particle positions uniquely: 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. Determinism can be restored by regarding the transformation $\mathbf{r}_i(t) \to \mathbf{r}_i(t) + \mathbf{f}(t)$ as a gauge transformation.

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

$$\mathbf{p}_i := \frac{\partial L}{\partial \dot{\mathbf{r}}_i} = \frac{m_i}{M}\sum_k m_k(\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_k) = m_i\dot{\mathbf{r}}_i - \frac{m_i}{M}\sum_k m_k \dot{\mathbf{r}}_k. \tag{12}$$

[14] For details, see Earman (1989, chapter 2, section 3).

These momenta are not independent but must satisfy three primary constraints, which require the vanishing of the x, y, and z components of the total momentum:

$$\phi_\alpha = \sum_i p_i^\alpha = 0, \qquad \alpha = 1, 2, 3. \tag{13}$$

These primary constraints are the only constraints – there are no secondary constraints – and they are first class. These constraints generate in each configuration variable $\mathbf{r}_i$ the same gauge freedom; namely, a Euclidean shift given by the same arbitrary function of time. The gauge-invariant quantities include such things as relative particle positions and relative particle velocities.

In this toy example there is simple connection between the senses of gauge freedom derived from the Lagrangian and Hamiltonian approaches. In more complicated cases, however, the connection between the two approaches is not so transparent.

5 The General Theory of Relativity as a gauge theory

The Hilbert action for the source-free Einstein gravitational field equations reads $\int_{\mathcal{M}} R\sqrt{-g}\, d^4x$, where R is the Ricci scalar and $g := \det(g_{ij})$, g_{ij} being the coordinate components of the spacetime metric. The diffeomorphism group $\mathit{diff}(\mathcal{M})$ of the spacetime manifold $\mathcal{M}$, which contains arbitrary functions of the independent variables in the action (here the spacetime coordinates x^i, $i = 1, 2, 3, 4$), is a variational symmetry. Thus, Noether's second theorem applies, telling us that we have a case of underdetermination. We expect that in the Hamiltonian formulation there will be constraints that generate non-trivial gauge. Our expectations are not disappointed. The configuration variables (the qs) are Riemann 3-metrics, interpreted as giving the intrinsic geometry of a 3-manifold that is to be embedded as a time slice of spacetime, and the conjugate momentum variables (the ps) are tensor fields related to the exterior curvature of the 3-manifold. When the crank of the Dirac algorithm is turned, it is found that the constraints are all primary first class. That sounds nice, but in fact when the details are unpacked three surprises/puzzles are revealed.

There are two families of constraints: the momentum constraints and the Hamiltonian constraints.[15] When the Poisson bracket algebra of these constraints is computed, it is found that it does not close, so this algebra is not a Lie algebra. This means that a defining feature of Yang–Mills theories is missing from the most natural formulation of GTR as a Hamiltonian 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

[15] 'Families' because for each point of space there is a momentum constraint and a Hamiltonian constraint.

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.

A second puzzle is how to find, among the Dirac gauge transformations on the Hamiltonian phase space of GTR, a counterpart of the action of the diffeomorphism group of spacetime. The most obvious way to identify a counterpart by finding a homomorphic copy of the Lie algebra of $\mathit{diff}(\mathcal{M})$ in the constraint algebra is blocked by the fact that the latter isn't a Lie algebra. One resolution of the puzzle is given by Isham and Kuchař (1986a; 1986b) who show that if the embedding variables (which describe how a 3-manifold is embedded as an initial value hypersurface of spacetime) and their conjugate momenta variables are adjoined to the phase space of GTR, there is a natural homomorphism of the Lie algebra of spacetime diffeomorphisms into the Poisson bracket algebra of constraints on the extended phase space. An alternative approach is taken by Ashtekar and Bombelli (1991), who show that Hamiltonian mechanics for general relativity does not require a $(3 + 1)$-cotangent bundle structure. Instead of taking the phase space of the theory to be the space $\Gamma(q, p)$ of instantaneous states, they work with the space $\hat{\Gamma}$ of entire histories or solutions to the Einstein field equations, which implies that dynamics is implemented not by a mapping from one state to another state in the same solution but as a mapping from one solution to another solution. The space $\hat{\Gamma}$ has a presymplectic structure given by a degenerate 2-form $\hat{\omega}$. There is no constraint surface, as in the $(3 + 1)$ formulation; rather, the gauge directions Y are given directly by the null vectors of $\hat{\omega}$. It turns out that two solutions lie on the same gauge orbit (i.e. integral curve of the gauge field Y) iff they are diffeomorphically related.

The third and most contentious puzzle arises from the fact that since the Hamiltonian constraints generate the motion, motion is pure gauge, and the observables of the theory are constants of the motion in the sense that they are constants along the gauge orbits. Taken at face value, the gauge interpretation of GTR implies a *truly* frozen universe: not just the 'block universe' that philosophers endlessly carp about – that is, a universe stripped of A-series change or shifting 'nowness' – but a universe stripped of its B-series change in that no genuine physical magnitude (= gauge-invariant quantity) changes its value with time. Philosophers of science have generally ignored this puzzle. But it deserves a resolution, either by showing how the 'no change' consequence can be avoided or by showing how the consequence can be reconciled with our perceptions of a world filled with (B-series) change. Since I have given my own resolution at length elsewhere, I will not repeat it here (see Earman, 2002). But lest one think that the problem depends on peculiarities of infinite-dimensional systems or of general relativity, I will mention that an analogue arises for a subclass of the humble theories discussed in the preceding sections. These are so-called reparameterization invariant Lagrangian theories where the action is invariant under $t \to f(t)$ for arbitrary $f(t)$. For such theories the canonical Hamiltonian, if non-zero, is a first-class constraint so that motion is pure gauge.

Finally, I will mention that the analysis of gauge recommended here helps to illuminate some of the issues surrounding the never-ending debate about the nature and status of the requirement of general covariance. What makes the issues so difficult to disentangle is that as the debate unfolded, every confusion it was possible to make was in fact made – coordinate transformations were confused with point transformations, relativity principles were confused with gauge principles, etc.[16] This is not the place to attempt a disentanglement, and I will have to be content with noting that much of the confusion is swept away by distinguishing two forms of the requirement of general covariance. The *weak requirement* demands that the equations of motion of a theory are valid in an arbitrary coordinate system (or, equivalently, that the equations should be covariant under an arbitrary coordinate transformation). Assuming that nature can be fully described by geometric object fields, this requirement is a restriction on the form rather than the content of a theory. The *strong requirement* demands that the spacetime diffeomorphism group is a gauge group of the theory. If the recommended Hamiltonian constraint apparatus is used to detect gauge freedom, then it is obvious that a theory can satisfy the weak requirement without at the same time satisfying the strong requirement. However, one can wonder whether the strong requirement, like its weak sister, is also a matter of form rather than content. That is, can a theory that satisfies the weak but not the strong requirement always be rewritten in a form that conforms to the latter? The application of the constraint formalism reveals that monkeying with a theory so as to make it satisfy strong general covariance may change the constraint structure of the theory and, thus, what counts as an observable (= gauge-invariant quantity). Arguably, such a change amounts to a change in the content rather than the form of the theory. It then becomes an empirical question as to whether nature is best described in terms of the observables of the original or of the new theory.[17]

[16] For a historical review of the debate, see Norton (1993).
[17] See Earman, 'Once more general covariance', unpublished manuscript.

6 The fibre bundle apparatus

How does the relevant fibre bundle structure arise for theories that are not presented in bundle-theoretic language? The constrained Hamiltonian formalism provides the basis of an answer. One would like it to be the case that when the reduced phase space $\tilde{\Gamma}$ of a constrained Hamiltonian system is formed by quotienting the constraint surface $\mathcal{C}$ by the gauge orbits, these orbits are the fibres of a bundle with base space $\tilde{\Gamma}$. 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 our desire is satisfied in that $\mathcal{C}/\mathcal{G}$ will be the base space of a $\mathcal{G}$-bundle whose fibres are the orbits of the group action. In practice, however, some of the stated suppositions may fail. For instance, although $\tilde{\Gamma}$ always exists as a topological space, it may lack a manifold structure – GTR is a particular example.[18] 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.

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, as we will see in section 8.

7 The reach of the constraint apparatus

In order to produce simple and understandable examples I have emphasized applications where, from the Lagrangian point of view, the gauge transformations are purely transformations of the independent variables[19] of the action and where these variables are identified with spatiotemporal variables. But nothing in the constraint formalism depends on these simplifying assumptions, and the formalism serves to identify the gauge freedom even when these assumptions do not hold.

However, there is one obvious and absolute limitation of the apparatus: it does not apply to equations of motion that are not derivable from an action principle. This might seem to be a mild limitation because almost all of the candidates for fundamental equations of motion in modern physics are derivable from an action principle. But appearances can be deceptive; in particular, the apparent ubiquity of equations of motion that are derivable from an action principle might represent a selection effect deriving from the facts that modern physicists always have

[18] In this case, however, the reduced phase space is not badly behaved since it is the disjoint union of manifolds; see Fischer and Moncrief (1996).

[19] These are sometimes referred to as 'gauge transformations of the first kind', whereas transformations depending only on the dependent variables are referred to as 'gauge transformations of the second kind'.

quantization in the back of their minds and that the standard cookbook procedures for producing a quantization start from a Hamiltonian formulation.

To get a feel for just how strong the limitation really is we would need to know what the necessary and sufficient conditions are for equations of motion to be derivable from an action principle. A special case of this problem is what is known as the *Helmholtz problem*: Consider a system of Newtonian second-order ordinary differential equations: $\Delta^i = 0$, $\Delta^i := \ddot{q}^i - F^i(q^i, \dot{q}^i, t)$; under what conditions does there exist a Lagrangian $L(q^i, \dot{q}^i, t)$ and a non-singular matrix A^{ik} such that $\Delta^i = A^{ik} EL_k$, where $EL_k = 0$ are the EL equations of L? Helmholtz found a set of necessary conditions which were later proved to be sufficient as well. Darboux proved that for $n = 1$ the Helmholtz conditions can always be satisfied. The case of $n = 2$ has also been solved. But the general problem remains unsolved (see, for example, Sarlet, 1982). Even less is known when more complicated equations of motion are considered and when singular Lagrangians are permitted.

The issue under discussion is also intimately linked with the status of determinism. Determinism becomes a trivial doctrine if whenever cracks appear in the doctrine we stand ready to paper them over by seeking gauge freedom. 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 whose equations of motion are derivable from an action principle, and suppose that the EL equations 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, as in the case of GTR or time reparameterization theories in general.) Thus, to decide just how *a priori* or contingent determinism is, it is crucial to know how strong is the demand that the equations of motion admit a (possibly singular) Lagrangian formulation – the stronger (respectively, weaker) the demand is, the more contingent (respectively, *a priori*) determinism is.

In the absence of any convincing argument to the effect that acceptable equations of motion must be derivable from an action principle, it seems necessary to confront the issue of how to identify gauge freedom and observables for equations of motion that are not derivable from an action principle. I know of no systematic approach to this issue.

Finally, I want to indicate a way in which the conviction that theories containing gauge freedom are described by constraints can be challenged. Independently of the desire to save determinism, a motivation for seeing gauge freedom at work comes from considerations of observability. Suppose, for example, that one has reasons for thinking that a complex-valued scalar field φ is not observable whereas combinations such as $\varphi^*\varphi$ are. If the action is constructed from such combinations it will be invariant under the group of transformations of the form $\varphi \to \varphi' = \exp(i\alpha)\varphi$, $\partial_\mu \alpha = 0$, which might be taken as gauge transformations even though the constraint apparatus does not apply since the group parameters do not involve arbitrary functions of the independent variables. In the case of quantum field theory this line on gauge leads to what I find to be unacceptable consequences, such as that unitarily inequivalent representations of the algebra of field operators are gauge equivalent and, thus, are to be regarded as merely different ways of describing the same physical situation (see Earman, this volume, Part III). And apart from considerations of quantum field theory, the present motivation for seeing gauge freedom is potentially confused. That the complex-valued scalar field φ is not observable or measurable is a necessary but not sufficient condition for consigning φ to the category of quantities which are not to be regarded as genuine physical magnitudes because their values can be changed without changing the real physical situation. Neverthless, it is important to recognize that there can be a number of different motivations for seeing gauge at work.

8 The quantization of gauge theories

There are at least four extant approaches to quantization of gauge theories.[20] The first is *gauge fixing*: 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 may fail to define a global transversal in the constraint surface, i.e. a hypersurface that meets each of the gauge orbits exactly once.

A second approach is *reduced phase space quantization*. Quotient out the gauge orbits to produce the reduced phase space. If this procedure goes smoothly (see section 10 below) the normal method of quantization can be applied to the resulting unconstrained Hamiltonian system. This approach faces the practical difficulty of having to solve the constraints, and even if one overcomes this difficulty one may find that the reduced phase space has features that complicate the quantization (see section 10 below).

The third approach is called *Dirac constraint quantization*. Here the procedure is to promote the first-class constraints to operators on a Hilbert space and then require that the vectors in the physical sector of this Hilbert space be annihilated by the constraint operators. Of course, the forming of the constraint operators is subject

[20] The standard reference on this topic is Henneaux and Teitelboim (1992).

to operator ordering ambiguities. But even modulo such ambiguities, it can happen that the resulting Dirac quantization is inequivalent to that obtained by reduced phase space quantization. In such a case, which is the correct quantization? And how would one tell? I will return to these matters below.

The fourth approach is called *BRST quantization* after Becchi, Rouet, Stora, and Tyutin. The idea is to mirror the original gauge symmetry by a symmetry transformation on an extended phase space obtained by adding auxiliary variables. The additional phase space variables are chosen so that the BRST symmetry has a simple form that facilitates quantization.

Apart from technical issues, the point to keep firmly in mind is that it is, presumably, the observables – in the sense of gauge-invariant quantities – of a constrained system that get promoted to quantum observables – in the sense of self-adjoint operators on the appropriate Hilbert space. Thus, it is hardly surprising that one of the key issues in the search for a quantum theory of gravity is what to take as the appropriate set of observables of classical GTR (see Isham, 1992).

9 The magical gauge argument

In the physics literature there is something called the 'gauge argument' that goes like this. Start with a free field which admits a 'global symmetry' and obeys (by Noether's first theorem) 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 an important strand of the logic of nature.

I am in agreement with Martin (2002a,b; this volume) who finds the 'getting something from nothing' character of the gauge argument too good to be true. In particular, a careful look at applications of this argument reveals that a unique theory of the interacting field results only if some meaty restrictions on the form of the final Lagrangian are implicitly in operation; and furthermore, the kind of locality needed for the move from the 'global symmetry' (invoking Noether's first theorem) to the 'local symmetry' (invoking Noether's second theorem) is not justified by an appeal to the no-action-at-a-distance sense of locality supported by relativity theory. Not only is there no magic to be found in the gauge argument, but the 'gauge principle' that prescribes a move from global to local symmetries for interacting fields can be viewed as output rather than input: for example, it can be viewed as the product of a self-consistency requirement (see, for example, Wald, 1986) or as a consequence of the requirement of renormalizability (see, for example, Weinberg, 1974).

What is missing is an explanation of the heuristic power of the gauge argument. That is a task someone else will have to perform.

10 Why gauge theories?

Since the presence of gauge freedom in a theory means that the theory employs quantities that lead to redundant descriptions in the form of a many–one correspondence between the state descriptions of the theory and the intrinsic physical state, why shouldn't the theory be purged of its 'surplus structure' (to use Michael Redhead's terminology) so as to achieve a one–one correspondence between state descriptions and physical states? I will now consider a series of possible responses.

10.1 Obstructions to getting an unconstrained Hamiltonian system

At least two kinds of obstructions can occur. (i) Quotienting out the gauge orbits may not produce a manifold. This situation occurs for GTR, but here the singularities are isolated and the reduced phase space is a disjoint union of manifolds (see Fischer and Moncrief, 1996). (ii) The reduced phase space is a manifold but this manifold is not the cotangent space of a reduced configuration space.

10.2 Ambiguities in quantization

Suppose now that no obstructions are encountered in passing to a reduced Hamiltonian phase space. But suppose that reduced phase space quantization gives a result that is physically inequivalent to Dirac constraint quantization, even allowing for operator ordering ambiguities. And suppose that the latter proves to be empirically correct. This is certainly a reason not to gauge out. But it also seems to be a reason to say that the 'gauge transformations' are not really gauge transformations, for it seems that relevant information is lost in passing to the reduced phase space. Such examples, however, may be fanciful. Dirac and reduced phase space quantization will coincide when the gauge group is unimodular;[21] and when the gauge group is not unimodular and the Dirac and reduced phase space quantizations are at odds, arguably the Dirac procedure is incorrect and a modified Dirac prescription is needed (see Duval *et al.*, 1991).

10.3 Future extensions and modifications of the theory

A gauge theory could be retained in order to preserve the possibility that a later development will provide a physically motivated way of breaking the original

[21] A group is unimodular if it carries a bi-metric volume.

gauge symmetry. This consideration has force to the extent that there are historical examples of such gauge breaking. However, such examples seem to be in short supply. The Aharonov–Bohm effect is sometimes touted as a relevant example, but it is better to take the moral of this example to be that gauge-invariant quantities must include non-local quantities such as the integral of the 4-potential around a closed loop.

10.4 Convenience

Here it is helpful to contrast various cases. (i) The use of the electromagnetic potentials in formulating Maxwell's theory seems to be purely a matter of convenience since Maxwell's equations can be stated and solved in any given application without use or mention of the potentials. (ii) More than mere convenience is involved in operating with gauge-dependent quantities in the toy example from section 4 of particle mechanics in Maxwellian spacetime. One can certainly work with gauge-independent inter-particle quantities; but for more than three particles, relative particle distances give an over-complete set of configuration variables for the reduced configuration space (see Belot, 2002), and writing unconstrained equations of motion requires an undemocratic choice of among these quantities. (iii) In our present state of knowledge, the use of gauge-dependent quantities in GTR – tensor fields on a manifold – cannot be ascribed to convenience or the desire to avoid an undemocratic choice. For at present we do not know how to do the mathematics of GTR purely in terms of the gauge-independent quantities; in particular, the usual way of working with differential equations is not an option since apparently the spacetime manifold disappears in the gauge-independent description, whatever exactly that may turn out to be (see section 5). Thus, at present it seems that treating GTR as a gauge theory is closer to *force majeure* than to convenience.

In sum, there is no one simple answer to 'Why gauge?' The answer will vary from case to case, and it can range from 'Because it makes the life of the physicist easier' to 'We don't seem have a choice in the matter.' But if the above is a fair summary, then the verdict must be that we don't presently have a satisfying answer to 'Why gauge?' since most of the reasons essayed, save for convenience, were of the in-principle but not-actually-in-practice form.

11 Conclusion

Nothing that I have said above is news to physicists. But it seems to me worth saying to a philosophical audience. Indeed, when I try to talk about constrained Hamiltonian systems to even my more knowledgeable colleagues in the philosophy of physics, they look at me as if I were speaking a Martian dialect. I certainly do not want to claim that the constraint formalism constitutes *the correct* approach to gauge; indeed, I doubt that there is an approach answering to that description, for gauge is such a broad and variegated concept that its explication requires the services of many different approaches and formalisms. But I do want to claim for the constraint formalism a number of virtues. In particular, among the extant approaches of which I am aware, it is the one with the broadest scope; for philosophers of science it has the advantage of immediately connecting to fundamental foundations issues, such as the nature of observables and the status of determinism; it explains how and under what conditions a fibre bundle structure emerges for theories which do not wear their bundle structure on their sleeves; and it calls attention to problems which arise in attempting to quantize gauge theories.

Independently of the merits of the constraint formalism, I want to urge that in getting a grip on the gauge concept, philosophers initially eschew the glitz of elementary particle physics, Yang–Mills theories, fibre bundles, etc., and concentrate instead on humbler examples. These examples often make it easier to see important conceptual connections, and they bring out the fact that the gauge concept is important for understanding not only Yang–Mills theories and theories of elementary particle physics, but Newtonian and classical relativistic theories as well.

Acknowledgements

I am grateful to Gordon Belot, Jeremy Butterfield, Katherine Brading, and Chris Martin for helpful comments and suggestions. Special thanks are due to Josep Pons for generously sharing his expertise on constrained Hamiltonian systems.

References

Ashtekar, A., and Bombelli, L. (1991). 'The covariant phase space of asymptotically flat gravitational fields'. In *Mechanics, Analysis and Geometry: 200 Years After Lagrange*, ed. M. Francaviglia, pp. 417–50. Amsterdam: Elsevier Science Publishers.

Batlle, C., Gomis, J., Pons, J. M., and Roman-Roy, N. (1986). 'Equivalence between the Lagrangian and Hamiltonian formalism for constrained systems'. *Journal of Mathematical Physics*, **27**, 2953–62.

Belot, G. (2002). 'Symmetry and gauge freedom'. *Studies in History and Philosophy of Modern Physics*, **34**, 189–225.

Bergmann, P. G. (1949). 'Non-linear field theories'. *Physical Review*, **75**, 680–5.

Bergmann, P. G., and Brunings, J. H. M. (1949). 'Non-linear field theories. II: Canonical equations and quantization'. *Reviews of Modern Physics*, **21**, 480–7.

Brading, K. A. (2002). 'Which symmetry? Noether, Weyl, and conservation of electric charge'. *Studies in History and Philosophy of Modern Physics*, **33**, 3–22.

Byers, N. (1999). ‘E. Noether’s discovery of the deep connection between symmetries and conservation laws’. *Israel Mathematical Society Proceedings*, **12**, 67–79.
Cabo, A., and Louis-Martinez, D. (1990). ‘On Dirac’s conjecture for Hamiltonian systems with first- and second-class constraints’. *Physical Review D*, **42**, 2726–35.
Dirac, M. (1950). ‘Generalized Hamiltonian dynamics’. *Canadian Journal of Mathematics*, **2**, 129–48.
(1951). ‘The Hamiltonian form of field dynamics’. *Canadian Journal of Mathematics*, **3**, 1–23.
(1964). *Lectures on Quantum Mechanics*. New York: Belfer Graduate School of Science Monographs Series.
Duval, C., Elhadad, J., Gotay, M. J., Śniatycki, J., and Tuynman, G. M. (1991). ‘Quantization and bosonic BRST theory’. *Annals of Physics*, **206**, 1–26.
Earman, J. (1989). *World Enough and Spacetime: Absolute vs. Relational Theories of Space and Time*. Cambridge, MA: MIT Press.
(2002). ‘Thoroughly modern McTaggart: or what McTaggart would have said if he had learned Einstein’s general theory of relativity’, *Philosophers’ Imprint 2*. Available online at http://www.philosophersimprint.org.
Fischer, A. E., and Moncrief, V. (1996). ‘A method of reduction of Einstein’s equations of evolution and a natural symplectic structure on the space of gravitational degrees of freedom’. *General Relativity and Gravitation*, **28**, 207–19.
Gomis, J., Henneaux, M., and Pons, J. M. (1990). ‘Existence theorem for gauge symmetries in Hamiltonian constrained systems’. *Classical and Quantum Gravity*, **7**, 1089–96.
Gotay, M. J., and Nester, J. M. (1984). ‘Apartheid in the Dirac theory of constraints’. *Journal of Physics A*, **17**, 3063–6.
Gràcia, X., and Pons, J. M. (1988). ‘Gauge generators, Dirac’s conjecture, and degrees of freedom for constrained systems’. *Annals of Physics*, **187**, 355–68.
Henneaux, M., and Teitelboim, C. (1992). *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press.
Isham, C. J. (1992). ‘Canonical quantum gravity and the problem of time’. In *Integrable Systems, Quantum Groups, and Quantum Field Theories*, ed. L. A. Ibot and M. A. Rodríguez, pp. 57–287. Boston, MA: Kluwer Academic.
Isham, C. J., and Kuchař, K. (1986a). ‘Representations of spacetime diffeomorphisms. I: Canonical parametrized field theories’. *Annals of Physics*, **164**, 316–33.
(1986b). ‘Representations of spacetime diffeomorphisms. II: Canonical geometrodynamics’. *Annals of Physics*, **164**, 288–315.
Kastrup, H. A. (1987). ‘The contributions of Emmy Noether, Felix Klein and Sophus Lie to the modern concept of symmetries in physical systems’. In *Symmetries in Physics (1600–1980)*, ed. M. G. Doncel. Barcelona: Universitat Autonoma Barcelona.
Lee, J., and Wald, R. M. (1990). ‘Local symmetries and constraints’. *Journal of Mathematical Physics*, **31**, 725–43.
Martin, C. (2002a). ‘Gauge principles, gauge arguments and the logic of nature’. *Philosophy of Science*, **69**, S221–34.
(2002b). ‘Gauging gauge: remarks on the conceptual foundations of gauge theory’. Ph.D. dissertation, University of Pittsburgh.
Maxwell, J. C. (1877). *Matter and Motion*. Reprinted (1951) New York: Dover.
Noether, E. (1918). ‘Invariante Variationsprobleme’. In *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse*, Book 2, 235–57.

Norton, J. D. (1993). ‘General covariance and the foundations of general relativity: eight decades of dispute’. *Reports on Progress in Physics*, **56**, 791–858.
Olver, P. J. (1993). *Applications of Lie Groups to Differential Equations*, 2nd edn. New York: Springer-Verlag.
Sarlet, W. (1982). ‘The Helmholtz conditions revisited: a new approach to the inverse problem of Lagrangian dynamics’. *Journal of Physics A*, **15**, 1503–17.
Wald, R. M. (1986). ‘Spin-two fields and general covariance’. *Physical Review D*, **33**, 3613–25.
Weinberg, S. (1974). ‘Recent progress in gauge theories of the weak, electromagnetic and strong interactions’. *Reviews of Modern Physics*, **46**, 255–77.
Wipf, A. (1994). ‘Hamilton’s formalism for systems with constraints’. In *Canonical Gravity: From Classical to Quantum*, ed. J. Ehlers and H. Friedrich, pp. 22–58. New York: Springer-Verlag.