7

From metaphysics to physics

GORDON BELOT AND JOHN EARMAN

1. Introduction

Michael Redhead began his Tarner Lectures by allowing that 'many physicists would dismiss the sort of question that philosophers of physics tackle as irrelevant to what they see themselves as doing' (1995, p. 1). He argued that, on the contrary, philosophy has much to offer physics: presenting examples and arguments from many parts of physics and philosophy, he led his audience towards his ultimate conclusion that physics and metaphysics enjoy a symbiotic relationship.

By way of tribute to Michael we would like to undertake a related project: convincing philosophers of physics themselves that the philosophy of space and time has something to offer contemporary physics. We are going to discuss the relationship between the interpretative problems of quantum gravity, and those of general relativity. We will argue that classical and quantum theories of gravity resuscitate venerable philosophical questions about the nature of space, time, and change; and that the resolution of some of the difficulties facing physicists working on quantum theories of gravity appears to require philosophical as well as scientific creativity. These problems have received little attention from philosophers. Indeed, scant attention has been paid to recent attempts to quantize gravity. As a result, most philosophers have been unaware of the problem of time in quantum gravity, and its relationship to the knot of philosophical and technical problems surrounding the general covariance of general relativity – so that it has been all too easy to dismiss this latter set of problems as philosophical contrivances. Consequently, philosophical discussion of space and time has suffered.

This point is best illustrated by attending to the contrast between what philosophers and physicists have to say about the significance of Einstein's hole argument. A version of this argument was used by Earman & Norton

(1987) to argue that it is a consequence of general covariance that substantivalism about the spacetime of general relativity can be maintained only at the price of indeterminism. Philosophical responses to this version of the argument divide quite strikingly into two camps. On the one hand, there are those who criticize the argument on the grounds that it relies upon a naive approach to modality. These authors argue that the *prima facie* force of the argument evaporates once one understands the subtlety of the modal semantics of spacetime points.[1] We believe, but will not argue here, that this variety of response lacks a coherent and plausible motivation.

The second sort of response to the hole argument is more radical, and its popularity more telling as a measure of the insularity of contemporary philosophy of space and time. The modalists acknowledge that the hole argument has *some* value: it points up a strange fact about the modal semantics of spacetime theories. A more radical response is to deny that the hole argument has *anything* at all to teach us about the nature of spacetime.[2] This is often combined with a general pessimism concerning the present state of philosophical discussion of space and time. Thus, Rynasiewicz contrasts the present state of the debate with its glorious past:

> What is remarkable about the substantival-relational debate is that, although it engaged natural philosophers from the seventeenth century into the nineteenth century and continues to be debated in academic philosophy, interest in the controversy on the part of twentieth century physicists has waned over the generations to virtually nil. (1992, p. 588)

Meanwhile, Leeds questions the interest of interpretative work on general relativity by contrasting the hole argument literature with another genre of philosophy of physics:

> There is an oddity here, it seems to me: for surely the philosophers of physics who work on these problems are the same men and women who, in another mood, are fond of comparing quantum mechanics with GTR, as the paradigm case of a theory which cries out for interpretation with the paradigm case of a theory which does not. (1995, p. 428)

[1] This camp subdivides into two factions: those who attempt to derive an appropriately sophisticated modal semantics for spacetime from some general framework (Bartels 1996; Brighouse 1994; Butterfield 1989; and Maudlin 1990); and those who take the required semantics as a primitive (Hoefer 1996; Maidens 1993; and Stachel 1993).

[2] Again there are two factions. On the one hand, we have those who hold the argument is fallacious because determinism is a formal property of theories, independent of questions of interpretation (Leeds 1995 & Mundy 1992). On the other hand, we have those who claim that the hole argument turns upon a piece of philosophy of language (the inscrutability of reference), and has nothing to do with philosophy of physics (Liu 1996 & Rynasiewicz 1996).

Discussion of the hole argument is often taken to be the epitome of irrelevant philosophy of physics. It is held, implicitly or explicitly, that it is *obvious* that there is nothing to the argument, since no physicist would entertain for a minute the proposition that general relativity is an indeterministic theory. It is supposed to be something of an embarrassment that philosophers have wasted so much time on this argument – how do they expect physicists to take them seriously? Typically, partisans of this line of thought believe that the hole argument is predicated on some sort of simple mistake. Most spectacularly, it is claimed that it has nothing in particular to do with general relativity at all. Rather, it is an artifact of a certain misguided way of thinking about language, naively mistaken for a bit of philosophy of physics:

> Such permutation arguments have been exploited at length by W. V. Quine, Donald Davidson, and Hilary Putnam to argue that a hankering for absolute criteria of individuation leads to an inscrutability of reference. The hole argument is nothing more than an application of the same techniques to space-time theories. If it yields relationist or anti-realist conclusions, these are conclusions which apply globally to any ontology. The substantival-relational debate, however, was a local one over the status of space and time.[3]

It is further alleged that philosophy of space and time has been led yet further astray by the suggestion that the 'solution' to the hole argument is to be found in the furthest reaches of metaphysics:

> I think that issues about whether *this* spacetime point could in some other world have been over *there* are not really questions about the nature of spacetime points, or indeed about physics at all, they are questions about *situations* or *possible worlds* – philosophers' constructions so loosely connected with reality that we can consistently answer these questions in any way that we care to. And in fact it seems to me that this had begun to be the consensus about these questions until Earman seemed to breathe new life into them via the connection with determinism. (Leeds 1995, p. 436)

These philosophers of physics paint a bleak picture indeed of the current state of the philosophy of space and time: having long ago lost its relevance to physics, it has recently degenerated into the worst sort of confused and eminently *philosophical* discussion.

This is in sharp contrast to the interest in the substantival–relational debate which is expressed by some physicists:

[3] Rynasiewicz (1996), p. 305. See also pp. 243–44 of Liu (1996) and p. 84 of Maudlin (1989).

> I would like to argue that the problem of quantum gravity is an aspect of a much older problem, that of how to construct a physical theory which could be a theory of an entire universe and not just a portion of one. This problem has a long history. It was, I believe, the basic issue behind the criticisms of Newtonian mechanics by Leibniz, Berkeley, and Mach. (Smolin 1991, p. 230)

This remark is not atypical: many physicists who work on canonical quantum gravity believe that the substantival–relational debate is directly relevant to their research.[4] In fact, many physicists emphasize the importance of interpretative questions about general relativity – often motivated by the belief that differences of opinion about the technical and conceptual difficulties of quantum gravity can be traced to differences of opinion concerning the classical theory. Thus, Rovelli asserts that

> many discussions and disagreements on interpretational problems in the quantum domain (for instance the famous 'time issue') just reflect different but unexpressed interpretations of the *classical* theory. Thus, the subtleties raised by the attempts to quantize the theory force us to reconsider the problem of observability in the classical theory. (1991c, pp. 297–8)

Furthermore, far from dismissing the hole argument as a simple-minded mistake which is irrelevant to understanding general relativity, many physicists see it as providing crucial insight into the physical content of general relativity. Thus, Isham uses a version of the hole argument to motivate an important claim about the observables of classical and quantum gravity:

> the diffeomorphism group moves points around. Invariance under such an active group of transformations robs the individual points of $\mathcal{M}$ of any fundamental ontological significance . . . This is one aspect of the Einstein 'hole' argument that has featured in several recent expositions (Earman & Norton 1987; Stachel 1989). It is closely related to the question of what constitutes an *observable* in general relativity – a surprisingly contentious issue that has generated much debate over the years and which is of particular relevance to the problem of time in quantum gravity. In the present context, the natural objects are Diff($\mathcal{M}$)-invariant spacetime integrals . . . Thus the 'observables' of quantum gravity are intrinsically non-local. (Isham 1993, p. 170)

Most surprisingly, one can even find physicists grappling with issues about the transworld identification of spacetime points:

[4] This belief is much less common among string theorists.

> The basic principles of general relativity – as encompassed in the term 'the principle of general covariance' (and also 'principle of equivalence') – tell us that there is no natural way to identify the points of one space-time with corresponding spacetime points of another.[5]

In short, a survey of the literature on quantum gravity reveals a very different picture of the relevance of philosophical work on the nature of space and time from that which is current among philosophers of physics. We do not mean to suggest that physicists are universally enthusiastic about the substantival–relational debate in general, or about the hole argument in particular. Nor, of course, are all philosophers of physics ill-disposed towards these topics. What *is* true is that philosophers of physics have tended to be unaware of the extent of the interest which physicists take in these issues. Philosophy of physics has suffered as a result: the interpretative inter-relationship between classical and quantum gravity has been missed; and the interest of the questions surrounding the general covariance of general relativity has been underestimated.

Our purpose in this short essay, is to bring these shortcomings to the attention of philosophers of physics, and to begin to redress them by giving a brief outline of the relationship between the interpretative problems of classical and quantum gravity, as we understand it.[6] Our focus in this paper is the canonical approach to quantum gravity, in which general relativity is first cast in Hamiltonian form, and then quantized via the canonical procedure. Unfortunately, the Hamiltonian formulation of general relativity is not entirely straightforward. Rather than being a true Hamiltonian system, general relativity is a gauge theory – and a somewhat peculiar one at that. The first task, undertaken in section 2, is to describe this formalism, and how its peculiarities derive from the general covariance of the standard formulation of general relativity. We will also discuss the interpretative problems of general relativity *qua* gauge theory. In the following section, we will see that the problem of time in quantum gravity – surely one of the deepest conceptual problems facing contemporary physics – follows from the gauge invariance of general relativity. Along the way, we will attempt to explicate the relationship between the somewhat unfamiliar conceptual problems of canonical gravity, and familiar philosophical problems about the nature of space, time, and change.

[5] Penrose (1996), p. 591. Penrose is led to this conclusion by the same considerations that motivate Rovelli and Isham; see especially p. 586.

[6] We discuss these issues at greater length in a companion paper, Belot & Earman (1999).

2. General relativity as a gauge theory

A Hamiltonian system consists of a phase space equipped with a real-valued function, H, the *Hamiltonian*. The geometric structure of the phase space is such that the specification of the Hamiltonian determines a unique curve $t \to x(t)$, called a dynamical trajectory, through each point of the space (see figure 7.1). Ordinarily, one thinks of the points of phase space as representing the dynamically possible states of some classical physical system, of the Hamiltonian as encoding information about the energy of each of these states, and the dynamical trajectory through a given point of phase space as representing the unique dynamically possible past and future of the state represented by that point. Thus interpreted, a Hamiltonian system constitutes a complete and deterministic description of a classical system.

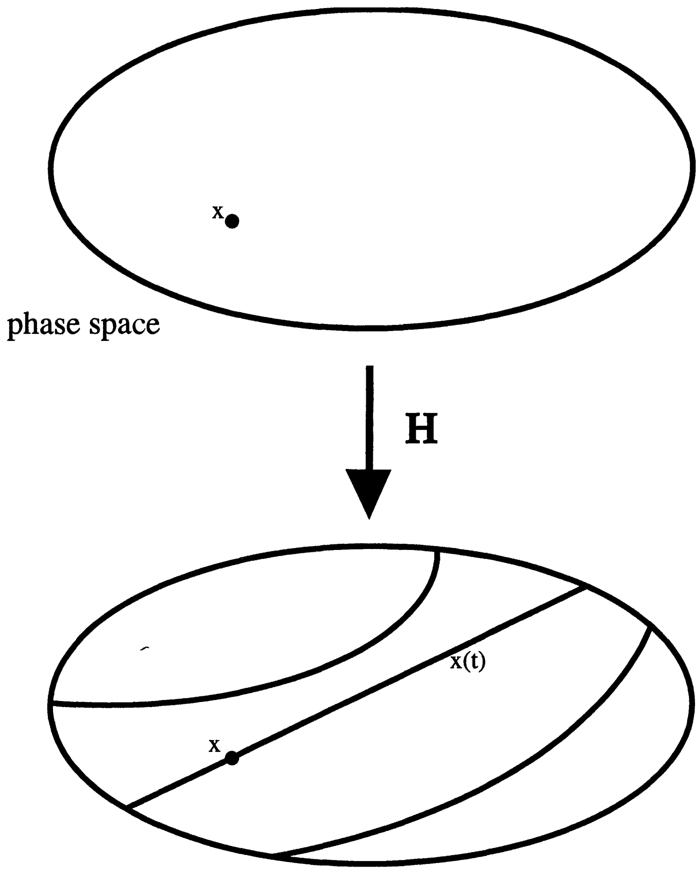

Figure 7.1 *Hamiltonian systems*

Unfortunately, the most natural formulations of many interesting classical theories – including electrodynamics and general relativity – are not strictly Hamiltonian. Rather they are gauge theories, in which the equations of motion fail to uniquely determine the evolution of the state in phase space. The geometric structure of the phase space of a gauge theory is somewhat weaker than that of a Hamiltonian system. For our purposes, the most important point is that the phase space of a gauge theory is naturally foliated by submanifolds of some fixed dimension, called gauge orbits (see figure 7.2). It is convenient to introduce the following notation and terminology: if x is a point in phase space, then $[x]$ is the unique gauge orbit in which x lies; if x and y are points of phase space then $x \sim y$ iff $[x]=[y]$; if f is a function on phase

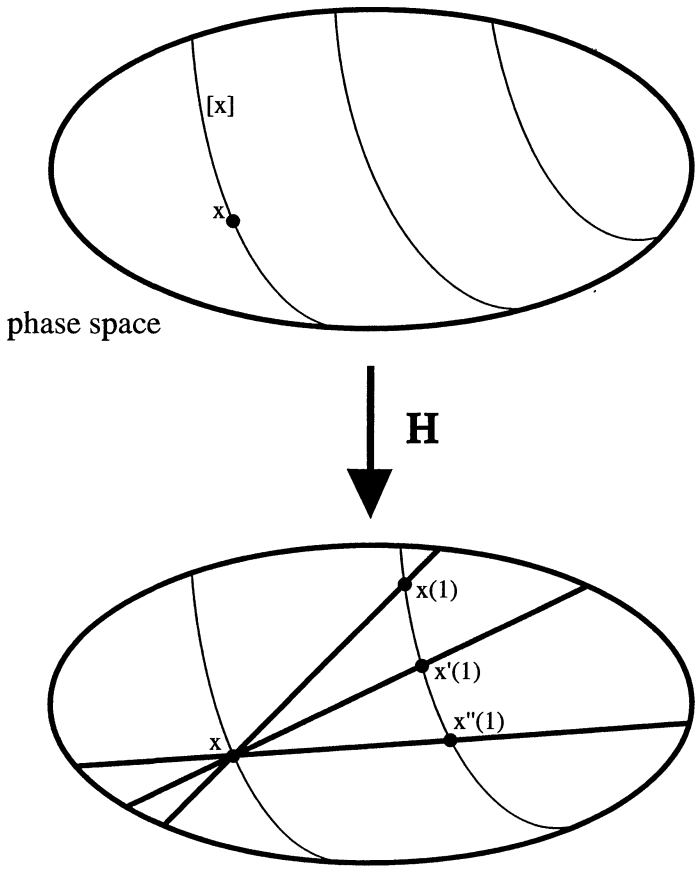

Figure 7.2 *Gauge systems.*

space which is constant on gauge orbits – i.e. if $x \sim y$ implies $f(x) = f(y)$ – then f is said to be *gauge invariant*. As in the Hamiltonian case, a gauge theory consists of a phase space equipped with a real-valued Hamiltonian H (we require that H be gauge invariant). But whereas in the former case the Hamiltonian together with the geometry of the phase space determined a unique dynamical trajectory through each point, in the gauge-theoretic case we find that there are infinitely many dynamical trajectories through each point of the phase space. Thus, if we fix an initial state, x, our theory is unable to tell us which point of phase space represents the state of the system at some later time t_1 – since we can find distinct dynamical trajectories, $x(t)$ and $x'(t)$, through our initial point x_0, and in general we expect that $x(t_1 \neq x'(t_1)$. What makes gauge theories interesting, however, is that although they are incapable of predicting which point of phase space represents the future state of the system, they *do* predict which gauge orbit that point will lie in – we find that $x(t_1) \sim x'(t_1)$ even if $x(t_1) \neq x'(t_1)$ (see figure 7.2).

The gauge freedom inherent in the equations of motion of a gauge theory complicates interpretation. It is possible to adopt the same *literal* approach which works so well for Hamiltonian systems, according to which each point of phase space corresponds to exactly one dynamically possible state. In this case, however, there will be physically real quantities which are not gauge invariant, since points lying in the same gauge orbit will correspond to distinct physically possible states of the system.[7] This straightforward approach has a serious disadvantage: it renders the theory indeterministic. Indeed, the state represented by our initial point, x_0, will have many possible futures: if $x(t)$ and $x'(t)$ are distinct dynamical trajectories through this point, then $x(t_1)$ and $x'(t_1)$ represent distinct physically possible future states of the system at time t_1. Note, however, that this indeterminism need not render the theory empirically inadequate: one can maintain that all *observable* quantities are gauge invariant, so that the theory can still be used to make determinate predictions of measurement outcomes, even if it cannot determine the evolution of all physical quantities.

Alternatively, we can require that our interpretation be *gauge invariant*, in the sense that all physically real quantities are represented by gauge invariant functions on phase space. In this case, points of phase space which lie in the same gauge orbit will correspond to the same dynamical state, and determinism will be rescued (since $x(t_1) \sim x'(t_1)$). *Ceteris paribus*, this sort

[7] Here and below we assume that two states are distinct if there is some physically real quantity which takes on different values in the two states.

of interpretation will be preferable to a literal interpretation, since we prefer to think of our classical theories as being deterministic.

All of this can be illustrated using the most familiar gauge theory, electrodynamics. Here the phase space is the set, $\{(A, E)\colon A, E\colon S \to \mathbb{R}^3, \text{div } E = 0\}$, of vector potentials and electric fields on physical space, S. The gauge orbits have the following structure: $(A, E) \sim (A', E')$, iff $E' = E$ and $A' = A + \text{grad } \Lambda$, for some $\Lambda\colon S \to \mathbb{R}$. The Hamiltonian for vacuum electrodynamics is $H = \int_S |E|^2 + |\text{curl } A|^2 dx^3$, and the equations of motion are $\dot{A} = -E$ and $\dot{E} = \text{curl}(\text{curl } A)$. These equations determine the evolution of E uniquely, but determine the evolution of A only up to the addition of the gradient of a scalar. That is, if $(A(t), E(t))$ and $(A'(t), E'(t))$ are two dynamical trajectories through the same initial point in phase space, then for all t we have that $E'(t) = E(t)$ and $A'(t) = A(t) + \text{grad } \Lambda(t)$, for some scalar Λ. If we give this theory a literal interpretation by stipulating that A corresponds to the velocity field of a material ether, then the theory becomes indeterministic – according to $A(t_1)$ *this* bit of ether ends up *here*, while according to $A'(t_1)$ it ends up *there*.[8] On the other hand, we can stipulate that along with the electric field, E, the only other physically real quantity is the magnetic field, $B \equiv \text{curl } A$. The theory is rendered deterministic, since B is a gauge invariant quantity whose evolution is determined uniquely by the equations of motion.[9] The vector potential becomes a mathematical fiction, and any information contained in A over and above that contained in B is, to borrow Redhead's apt phrase, *surplus structure* (see Redhead 1975).

We now turn to the more complicated case of general relativity. The points of the phase space of general relativity should represent instantaneous states of the gravitational field. Thus, it is natural to build the phase space out of points which represent the geometries of Cauchy surfaces of models of general relativity. We proceed as follows. We fix a three-dimensional manifold, Σ. We now imagine that this manifold is embedded in a model, (M, g), of general relativity as a Cauchy surface. What geometrical information does Σ inherit from (M, g)? The answer is: the first and second fundamental forms, h and k, of Σ considered as a submanifold of (M, g). Here h is just the Riemannian metric which results from restricting g to Σ, and k is, *very* roughly, the time derivative of h. Thus we can think of h and k as being the position and momentum variables of our gravitational field theory. The phase space of general relativity is just the set of

8 Such an interpretation would, presumably, be supplemented with an account of measurement which would imply that this indeterminism would be empirically undetectable.

9 See Belot (1998) for an account of the difficulties which this interpretation faces in light of the Aharonov–Bohm effect.

pairs, (h, k), which can arise from embedding Σ as a Cauchy surface of a model of general relativity. The gauge orbits of this phase space have a strikingly simple structure: (h, k) and (h', k') lie in the same gauge orbit iff they can be viewed as Cauchy surfaces of the *same* model of general relativity.

Thus, if we fix a model of general relativity and look at the geometries, (h, k) and (h, k'), corresponding to two distinct Cauchy surfaces, then these geometries will lie in the same gauge orbit of general relativity. This means that the dynamical trajectory which joins these two points must lie entirely within their common gauge orbit. It follows that in general relativity the Hamiltonian assumes a very simple form: $H \equiv 0$. This is, in fact, a straightforward consequence of the general covariance of the theory: in order to have a non-zero Hamiltonian, one must have access to a preferred parameterization of time (see chapter 4 of Henneaux & Teitelboim 1992 for a careful discussion). This feature, that the dynamical trajectories are restricted to gauge orbits, rather than passing from one orbit to another, distinguishes general relativity from other familiar gauge theories and is the source of some of the most interesting interpretative problems of classical and quantum gravity.

Indeed, once general relativity has been formulated as a gauge theory, we can reformulate the hole argument so that it depends upon the structure of the phase space of general relativity rather than upon the diffeomorphism invariance of the standard formulation of the theory. Above, we noted if $x(t)$ and $x'(t)$ are two dynamical trajectories passing through the same initial point, x_0, then these trajectories represent distinct dynamical futures for x_0 under a literal interpretation, but represent the same dynamical future under a gauge invariant interpretation. Thus, general relativity, like any gauge theory, is indeterministic under a literal interpretation, and deterministic under a gauge invariant interpretation. But notice that in general relativity $x(t)$ and $x'(t)$ correspond to the same four-dimensional geometry of spacetime (since they each represent a sequence of instantaneous geometries which belong to the same gauge orbit). This is the core of the hole argument, re-expressed in the language of gauge theories.

The connection with the substantival–relational debate can be recovered as follows. First, notice that in general relativity $x(t)$ and $x'(t)$ correspond to the same four-dimensional geometry. Next, consider the condition which Leibniz and Clarke agreed constituted a good criterion for distinguishing absolutists from relationalists: the absolutist will affirm, while the relationalist will deny, that there could be two worlds whose contents instantiated the same spatial relations, but which were numerically distinct in virtue of the fact that different points of space would be occupied by the material objects of the two worlds. In the context of general relativity, the natural generalization of this criterion is: substantivalists will affirm, while relationalists will deny, that there could be two general relativistic worlds which instantiated the same four dimensional geometry which were numerically distinct in virtue of the fact that the geometrical relations of these worlds would be differently shared out among the spacetime points of the worlds.[10] Thus, substantivalists will view $x(t)$ and $x'(t)$ as representing distinct instantiations of the given four-geometry by a set of existent spacetime points. Relationalists, on the other hand will maintain that all instantiations of a given four-dimensional geometry are numerically identical – $x(t)$ and $x'(t)$ correspond to the same physical possibility. Hence relationalism is a gauge invariant interpretation of general relativity.

At this point it would seem to be mandatory to adopt a gauge invariant interpretation of general relativity. Otherwise, we are committed to ruling general relativity to be indeterministic for the slimmest of reasons: a metaphysical preference for substantivalism. Certainly, there are a number of prominent gravitational physicists who accept this line of thought (see, e.g., Rovelli 1991c and 1997). Most philosophers who have written on the hole argument concur (although they are more likely to opt for some sophisticated form of substantivalism than for Rovelli's robust relationalism). In the next section, we will discuss the bearing that considerations arising out of quantum gravity have on this question. Our conclusion will be that, when the dust settles, these considerations may well override any grounds for settling the substantival–relational dispute which are internal to general relativity itself. Before turning to this argument, however, we would like to point out that, even at the classical level, the formulation of a cogent gauge invariant interpretation of general relativity is by no means a straightforward task.[11]

[10] For the purposes of this paper, we bracket the question of the cogency of the sort of substantivalists, mentioned briefly in 7.1, who deny the Leibniz–Clarke condition; see Earman (1989) and Belot (1999). For present purposes, it suffices to observe that *some* varieties of substantivalism are literal interpretations of general relativity.

[11] Many philosophers of physics have assumed that the formulation of a gauge-invariant interpretation of general relativity is an easy and attractive option – simply count diffeomorphic models as physically equivalent. We will see below, however, that this strategy runs into serious technical and conceptual difficulties when we attempt to formulate interpretations in terms of the phase space of the theory rather than individual models. The upshot for the substantival–relational debate remains unclear. But the existence of these problems, and the fact that they do not arise for other gauge theories such as electromagnetism, makes it wholly implausible that the issues surrounding the hole argument are pseudo-problems or merely a recapitulation of familiar problems about reference.

Our first worry is of a technical nature. Recall that the problem of isolating the gauge invariant quantities of a theory is closely related to the problem of formulating a gauge invariant formulation of that theory. Indeed, according to a gauge invariant interpretation all physically real quantities are represented by gauge invariant quantities on phase space. It is thus impossible to fully specify a gauge invariant interpretation until the gauge invariant functions have been identified. In the case of general relativity, this is quite a tall order. Very few gauge invariant quantities are known. Worse, in the spatially compact case it has been proven that there are no gauge invariant quantities which are local (i.e., which can be written as integrals over Σ of h, k, and a finite number of their derivatives; see Torre 1993). Thus, there is reason to worry that the gauge invariant quantities of general relativity may not be suitable candidates for the ontology of an interpretation of a classical field theory like general relativity.

Thus, it is not at all trivial to formulate a gauge invariant interpretation of general relativity. As long as the existence of a sufficient number of suitable gauge invariant quantities of general relativity remains an open question, a dark cloud hangs over the programme of giving a gauge invariant interpretation of the theory. We contend that an honest approach to the interpretative enterprise requires one to suspend judgement until these difficult technical questions are settled. In support of this contention, we note that the isolation of the gauge invariant quantities of the theory appears to be a prerequisite for a gauge invariant approach to quantization. And, we claim, an interpretation which supports quantization is deeper than one which does not. We conclude that the present state of ignorance about the existence of gauge invariant quantities for general relativity should give pause to advocates of gauge invariant interpretations.[12]

Our second problem is primarily conceptual in nature, and prefigures the problems to be discussed in the next section. Although it lies near the core of the conceptual difficulties facing attempts to quantize gravity, it is simple enough to state: *prima facie, gauge invariant interpretations of general relativity imply that time and change are illusions*. Lest it be thought that this is a pseudo-problem concocted by benighted philosophers, let us begin with a formulation from a leading gravitational physicist:

> How can changes in time be described in terms of objects which are completely time independent? In particular, since the only physical, and thus

[12] This is closely related to the claim, advanced in Earman (1989), that relationalists are obliged to produce formulations of physical theories which can be expressed in relationally pure vocabulary.

> measurable quantities are those which are time independent, how can we describe the rich set of time dependent observations we make of the world around us? (Unruh 1991, p. 266)

The argument here is straightforward. Fix a model of general relativity, (M, g), with two Cauchy surfaces, Σ_1 and Σ_2. If we accept that the only physically real quantities of general relativity are gauge invariant, then it follows that there is no physically real quantity which takes on different values when evaluated on Σ_1 and Σ_2. This is to say that there is no change in the world described by (M, g), since no physically real quantity evolves in time. *Prima facie*, proponents of gauge invariant interpretations of general relativity are committed to the view that change is illusory. This is a very radical thesis – it is, for instance, much stronger than the doctrine that there is no room for temporal becoming in a relativistic world, since proponents of the tenseless theory of time are confident that they can account for the existence of change (see, e.g., Mellor 1981).

There has been a great deal of discussion of this thesis, and its consequences, in the physics literature. It is clear that at the classical level it does not affect the routine business of applying general relativity – there is no difference of opinion as to the predictions of general relativity between *Parmenideans* who hold that time and change are illusory and the *Heraclitians* who believe that change is a fundamental reality. None the less, the question of the reality and nature of time and change is the subject of debate among physicists (see, e.g., the exchange between Kuchař and Rovelli on pp. 138–40 of Ashtekar & Stachel 1991). The reason for this is straightforward: it is felt that one must correctly understand the nature of time and change in the classical theory if one is to make any progress on the deep conceptual problems of quantum gravity.

3. Quantum gravity and the problem of time

In the previous section, we saw that it was possible to formulate general relativity as a gauge theory. This cast the interpretative problems of general relativity in a new light: it became clear that the hole argument is just a special case of the general observation that literal interpretations of gauge theories are indeterministic. This observation motivates us to search for gauge invariant interpretations of general relativity. But this turns out to be unexpectedly difficult: unsolved technical problems and daunting conceptual difficulties stand in our way. The latter, especially, are troubling:

adopting a gauge invariant interpretation of general relativity seems to require us to revise out most fundamental metaphysical categories. Thus we find ourselves in the following position: having noticed that general relativity is a gauge theory, we attempt to apply the interpretative strategy which works so well for other gauge theories, such as electrodynamics. But here the peculiar nature of general relativity *qua* gauge theory – the fact that the Hamiltonian is zero, so that dynamical trajectories are restricted to gauge orbits – forces us to confront difficulties that have no analogues in other familiar gauge theories. The case is similar when we attempt to quantize general relativity. Here the most obvious strategy is to apply to general relativity the algorithm of canonical quantization which works so well for other gauge theories. In this section we will see that this leads to tremendous conceptual difficulties. These may be traced back, via the vanishing of the classical Hamiltonian, to the general covariance of general relativity.

The algorithm for quantizing gauge theories is simple enough in outline.[13] As in ordinary quantum mechanics, one begins by selecting a set of classical position and momentum variables, and then constructing a representation of their algebra as an algebra of operators on the Hilbert space $L^2(Q, \mu)$ (here Q is the classical configuration space and μ is some appropriate measure on Q). One then isolates the subspace, $\mathcal{H}$, of $L^2(Q, \mu)$ consisting of gauge invariant wave functions. Once equipped with a suitable inner product, $\mathcal{H}$ will be the space of states for our quantum theory. In order to complete the construction, we need to introduce an algebra of gauge invariant quantum observables on $\mathcal{H}$, and a quantum Hamiltonian, $\hat{H}$, which determines the dynamics of the theory via the Schrödinger equation.

In the case of electrodynamics, we take the components of A and E at each point of physical space, S, to be our classical position and momentum variables. We then construct $L^2(Q, \mu)$ by taking Q to be the set, $\{A : S \rightarrow \mathbb{R}^3\}$, of vector potentials. The elements of $L^2(Q, \mu)$ are wave functions over Q: complex functionals of the form $\Psi[A]$. We construct $\mathcal{H}$ by restricting our attention to those $\Psi \in L^2(Q, \mu)$ which are gauge invariant in the sense that $A \sim A'$ implies $\Psi[A] = \Psi[A']$. We then find a set of self-adjoint operators on $\mathcal{H}$ which represent an algebra of gauge invariant classical observables, and impose a quantum version of the classical Hamiltonian, $H = \int_S |E|^2 + |\text{curl A}|^x 3$. The result is a quantum theory of

[13] The details, of course, involve many subtleties, which we gloss over in the following. It is safe to assume that the quantization of gravity faces all of the technical difficulties present in other quantum field theories – operator ordering ambiguities, anomalies, problems of regularization and renormalization, etc. – and then some.

the familiar type: a Hilbert space carrying a self-adjoint representation of an algebra of observables, with dynamics given by a Schrödinger equation.

It is possible, at least formally, to apply this algorithm to general relativity.[14] One takes the components of h and k at each point of Σ as the classical position and momentum variables. Thus, our states will be wavefunctions on $Q = \text{Riem}\ \Sigma$, the space of Riemannian metrics on Σ. We restrict our attention to gauge invariant wavefunctions.[15] The observables must remain unspecified for the time being, since so few classical gauge invariant quantities have been identified. But it is easy to write down the correct quantum Hamiltonian: $\hat{H} \equiv 0$. Thus, the Schrödinger equation becomes trivial:

$$\frac{\partial \Psi[h]}{\partial t} = i\hbar \hat{H} \Psi[h] = 0.$$

Since the quantum Hamiltonian is zero, there is no evolution in time of the quantum states. This is the core of the *problem of time*: there appears to be no time or change in quantum gravity. This is not surprising: the algorithm sketched above for the construction of quantum gravity treats the gauge invariance of general relativity in strict analogy with the gauge invariance of other theories. And we know from the discussion of section 2 that embracing an interpretation of general relativity which is gauge invariant in this sense involves, at least *prima facie*, renouncing the existence of time and change. But, whereas in the classical domain our interpretative beliefs did not interfere with our ability to apply the theory, it appears to be impossible to understand this gauge invariant theory of quantum gravity as a theory about *our* world, replete as it is with change. For, naively applying a fragment of the conceptual apparatus of ordinary quantum mechanics, it appears that the states, $\Psi[h]$, of our theory of quantum gravity tell us that the probability of measuring the spatial geometry to be (Σ, h) at a given instant is given by $|\Psi[h]|^2$. But according to our theory the state never evolves, and so the probability of obtaining a given three-geometry as the outcome of a measurement is constant in time. But this contradicts one of

[14] I.e. although technical problems – such as the prohibitive difficulty of constructing an appropriate measure on the configuration space of general relativity – prevent one from achieving a fully rigorous formulation of quantum gravity along these lines, it is possible to formally manipulate the equations. Even at this level, one can see that the resulting theory of quantum gravity would run into serious interpretative problems.

[15] Naively, we might hope that the gauge invariant wavefunctions would be those for which $\Psi[h] = \Psi[h']$ if h and h' can be viewed as spatial geometries of Cauchy surfaces of the same model. Unfortunately, the actual situation is considerably more complicated. See pp. 189–93 and 225–6 of Isham (1993).

the most basic tenets of modern cosmology: that the geometry of the universe is temporally evolving.

Thus, the most straightforward approach to the construction of a quantum theory of gravity results in a theory which appears to be incapable of describing our world. In recent years, many attempts have been made to remedy this situation: either by showing that this appearance is misleading and that one can indeed construct a viable theory of quantum gravity by treating the gauge invariance of general relativity in analogy with the gauge invariance of other theories; or by suggesting alternative routes to quantum gravity which rely upon unorthodox interpretations of the gauge invariance of general relativity.[16] We will briefly describe four such attempts, and sketch their interpretative underpinnings. Although we will not go into details, it is safe to assume that each of these proposals is fraught with serious technical and conceptual difficulties (see Belot & Earman 1999 for further details and references).

One of the most radical proposals is Barbour's timeless interpretation of quantum gravity (see Barbour 1994a, b). Barbour accepts the Parmenidean interpretation of general relativity, and the approach to quantum gravity sketched above to which it leads. He also endorses the account of measurement according to which the probability of finding the spatial geometry (Σ, h) is given by $|\Psi[h]|^2$. He must, therefore, face the full force of the problem of time. His response is to bite the bullet: Barbour acknowledges that he is committed to the view that the probability of measuring (Σ, h) does not change in time; his explanation for this surprising result is that *there is no time*. On his view, what exists is a single moment, and a wavefunction which tells us the probabilities for possible outcomes of measurements of the geometry of this instant. The geometries which are likely measurement outcomes are supposed to encode information which would make it appear *as if* the universe had a past and future. But in fact, on Barbour's view, all that exists, and all that we experience, is a single magical moment.

Rovelli has developed a somewhat less radical Parmenidean approach to quantum gravity (Rovelli 1991a, b). He begins by endorsing the Parmenidean approach to general relativity, and the relationalism which underlies it. Thus, he posits that the physically real quantities in general relativity are gauge invariant. These are constants of motion of the theory, in the sense that they take on the same value at any two Cauchy surfaces of a given model of general relativity. Thus, if we are talking about a system

[16] Kuchař (1992) is the canonical survey.

containing a rocket, 'the mass of the rocket' will not be a physically real quantity, since it changes over time. But quantities of the form 'the mass of the rocket at blast-off' and 'the mass of the rocket when it docks at the space-station' *will* be constants of motion – they take on a single value for each model of the theory. Rovelli's insight is that we can give the set of constants of motion of the theory an internal structure by grouping together constants of motion: we can form an 'evolving constant of motion' whose members are just the constants of motion which give the mass of the rocket at each instant. The resulting set will form a one-parameter family. We can even write down an equation which describes the change in the value of the evolving constant as the parameter is varied. Rovelli's hope is that we can do the same at the quantum level: group the quantum observables which correspond to classical constants of motion into quantum evolving constants. This is a technically daunting task. If it can be carried out, then one could hope to write down equations which would govern the evolution in parameter time of the expectation values of the quantum observables. One would then have explained how the appearance of time can arise out of fundamentally timeless structures. This would not, however, amount to 'finding time' in quantum gravity, since Rovelli doubts that there is a unique time hidden here. Rather, one expects that there will be many ways of constructing evolving constants, classical and quantum. We should be opportunistic about selecting a technique which suits the model at hand, and our psychological experience of time, without reading our pragmatic decision back into nature.

On the Heraclitian side, we again find a very radical approach which privileges spatial structure, and a less radical, but more ambitious proposal. The former consists of breaking the general covariance of general relativity by introducing a privileged time parameter.[17] Using this preferred time coordinate, we can rewrite general relativity as a time-dependent Hamiltonian system with a non-zero Hamiltonian. It is then in principle possible to construct a quantum field theory of gravity in which the privileged classical time parameter provides the background for the evolution of the quantum states. Thus, general relativity and quantum gravity are recast as theories of the evolution of the gravitational field in time. In particular, general relativity becomes a theory of the evolution in time of the geometry of space. One ends up with theories of gravity which are no more difficult

[17] This can be done in a number of ways. Popular approaches include the introduction of special forms of matter and the privileging of the foliations of spacetime by Cauchy surfaces of constant mean curvature. See Kuchař (1992) for references and discussion.

(or easy!) to interpret than other classical and quantum field theories. Of course, this is achieved at the price of sacrificing one of the conceptual cornerstones of contemporary physics: the idea that the spirit behind the general covariance of general relativity forbids one from introducing preferred coordinate systems.

Kuchař's internal time proposal represents a more plausible Heraclitian alternative (see Kuchař 1972 and 1993). He endeavours to respect the spirit of the general covariance of general relativity without treating it as a principle of gauge invariance. His starting point is the conviction that if one wants to make sense of our experience of change, then one must accept that there are physically real quantities of general relativity and quantum gravity which are not gauge invariant. As argued in section 2, such a position is closely associated with substantivalism about the spacetime of general relativity. Kuchař's goal is to isolate within the classical phase space some structure which deserves to be called temporal, but which would not single out a preferred time parameter.[18] This temporal structure would allow one to explicate a sense in which the physically real quantities of general relativity evolve, and hence cannot be gauge invariant. One would then use the same temporal structure as the background against which the states of quantum gravity would exhibit non-gauge invariant evolution. Kuchař's programme is very ambitious technically, and is, as yet, incomplete. But it suggests a way in which substantivalism, despite its shortcomings, can underwrite a distinctive and intriguing approach to the quantization of general relativity.

4. Conclusion

In the previous section, we saw how the problem of time in quantum gravity arises out of the conceptual difficulties surrounding the general covariance of general relativity. We sketched four proposed solutions to the problem of time, and saw how each was linked to a definite view about the nature of change and time in the classical and quantum world, and to a view about the nature of the spacetime of general relativity. One expects, of course, that each of these programmes, if developed rigorously, would lead to a different theory of quantum gravity. If one of them should turn out to be empirically adequate, that fact would have interpretative repercussions at the classical

[18] The vagueness here is, of course, our own: we gloss over the details of Kuchař's sophisticated and elegant construction.

level – if an interpretation of general relativity suggests a given approach to quantization, then one is bound to revise one's interpretative judgements should that approach prove untenable. Thus, we find the interpretative problems of general relativity and quantum gravity to be bound in a close relationship: we cannot settle one set of questions without this having repercussions for the other set. And so long as the way forward in quantum gravity is unclear, physicists will continue to ponder and to debate metaphysical questions about the nature and existence of spacetime and change. Thus, we reach the same conclusion as Michael did in his Tarner lectures: 'that physics and metaphysics blend into a seamless whole, each enriching the other, and that in very truth neither can progress without the other' (Redhead 1995, p. 87).

References

Ashtekar, A. & J. Stachel (eds) (1991). *Conceptual Problems of Quantum Gravity* (Boston: Birkhäuser).

Barbour, J. (1994a). 'The Timelessness of Quantum Gravity: I. The Evidence from the Classical Theory', *Classical and Quantum Gravity* **11**, 2853–73.

(1994b). 'The Timelessness of Quantum Gravity: II. The Appearance of Dynamics in Static Configurations', *Classical and Quantum Gravity* **11**, 2875–97.

Bartels, A. (1996). 'Modern Essentialism and the Problem of Individuation of Spacetime Points', *Erkenntnis* **45**, 25–43.

Belot, G. (1998). 'Understanding Electromagnetism', *British Journal for the Philosophy of Science* **49**, 531–55.

(1999). 'Rehabilitating Relationalism', *International Studies in Philosophy of Science* **13**, 35–52.

Belot, G. & J. Earman (1999). 'Pre-Socratic Quantum Gravity', forthcoming in C. Callender & N. Huggett (eds.) *Philosophy Meets Physics at the Planck-Scale* (Cambridge: Cambridge University Press).

Brighouse, C. (1994). 'Spacetime and Holes', in D. Hull, M. Forbes & R. M. Burian (eds.) *PSA 1994*, Vol. 1 (East Lansing, MI: Philosophy of Science Association), pp. 117–25.

Butterfield, J. (1989). 'The Hole Truth', *British Journal for the Philosophy of Science* **40**, 1–28.

Earman, J. (1989). *World Enough and Space-Time* (Cambridge, MA: MIT Press).

Earman, J. & J. Norton (1987). 'What Price Spacetime Substantivalism? The Hole Story', *British Journal for the Philosophy of Science* **38**, 515–25.

Henneaux, M. & C. Teitelboim (1992). *Quantization of Gauge Systems* (Princeton: Princeton University Press).

Hoefer, C. (1996). 'The Metaphysics of Space-Time Substantivalism', *Journal of Philosophy* **XCIII**, 5–27.

Isham, C. (1993). 'Canonical Quantum Gravity and the Problem of Time', in L. A. Ibort and M. A. Rodriguez (eds.) *Integrable Systems, Quantum Groups, and Quantum Field Theories* (Dordrecht: Kluwer, 1993), pp. 157–288.

Kuchař, K. (1972). 'A Bubble-Time Canonical Formalism for Geometrodynamics', *Journal of Mathematical Physics* **13**, 768–81.

(1992). 'Time and Interpretations of Quantum Gravity', in G. Kunsatter, D. Vincent and J. Williams (eds.) *Proceedings of the 4th Canadian Conference on General Relativity and Astrophysics* (Singapore: World Scientific, 1992), pp. 211–314.

(1993). 'Canonical Quantum Gravity', in R. J. Gleiser, C. N. Kozameh and O. M. Moreschi (eds.) *General Relativity and Gravitation 1992: Proceedings of the Thirteenth International Conference on General Relativity and Gravitation held at Cordoba, Argentina, 28 June–4 July 1992* (Philadelphia: Institute of Physics Publishing), pp. 119–50.

Leeds, S. (1995). 'Holes and Determinism: Another Look', *Philosophy of Science* **62**, 425–37.

Liu, C. (1996). 'Realism and Spacetime: Of Arguments Against Metaphysical Realism and Manifold Realism', *Philosophia Naturalis* **33**, 243–63.

Maidens, A. (1993). '*The Hole Argument: Substantivalism and Determinism in General Relativity*', Ph.D. thesis, University of Cambridge.

Maudlin, T. (1989). 'The Essence of Space-Time', in M. Forbes & A. Fine (eds) *PSA 1988*, Vol. 2 (East Lansing MI: Philosophy of Science Association), pp. 82–91.

(1990). 'Substances and Space-Time: What Aristotle Would Have Said to Einstein', *Studies in History and Philosophy of Science* **21**, 531–61.

Mellor, D.H. (1981). *Real Time* (Cambridge: Cambridge University Press).

Mundy, R. (1992). 'Space-Time and Isomorphism', in D. Hull, M. Forbes & K. Okruhlik (eds.) *PSA 1992*, Vol. 1 (East Lansing, MI: Philosophy of Science Association), pp. 515–27.

Penrose, R. (1996). 'On Gravity's Role in Quantum State Reduction', *General Relativity and Gravitation* **28**, 581–600.

Redhead, M. (1975). 'Symmetry in Intertheory Relations', *Synthese* **32**, 77–112.

(1995). *From Physics to Metaphysics* (Cambridge: Cambridge University Press).

Rovelli, C. (1991a). 'Is There Incompatibility Between the Way Time is Treated in General Relativity and in Standard Quantum Mechanics?', in A. Ashtekar & J. Stachel (eds.) *Conceptual Problems of Quantum Gravity* (Boston: Birkhäuser, 1991), pp. 126–40.

(1991b). 'Time in Quantum Gravity: An Hypothesis', *Physical Review D* 43, 442–56.

(1991c). 'What is Observable in Classical and Quantum Gravity', *Classical and Quantum Gravity* **8**, 297–316.

(1997). 'Halfway Through the Woods: Contemporary Research on Space and Time', in J. Earman & J. Norton (eds.) *The Cosmos of Science* (Pittsburgh: University of Pittsburgh Press, 1997), pp. 180–223.

Rynasiewicz, R. (1992). 'Rings, Holes and Substantivalism: On the Program of Leibniz Algebras', *Philosophy of Science* **59**, 572–89.

(1996). 'Absolute versus Relational Space-Time: An Outmoded Debate?', *Journal of Philosophy* **XCIII**, 279–306.

Smolin, L. (1991). 'Space and Time in the Quantum Universe', in A. Ashtekar & J. Stachel (eds.) *Conceptual Problems of Quantum Gravity* (Boston: Birkhäuser, 1991), pp. 228–91.

Stachel, J. (1989). 'Einstein's Search for General Covariance, 1912-1915', in D. Howard and J. Stachel (eds.) *Einstein and the History of General Relativity* (Boston: Birkhäuser, 1989), pp. 63–100.

(1993). 'The Meaning of General Covariance', in J. Earman, A. Janis, G. Massey and N. Rescher (eds.) *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolph Grünbaum* (Pittsburgh: University of Pittsburgh Press, 1993), pp. 129–60.

Torre, C. G. (1993). 'Gravitational Observables and Local Symmetries', *Physical Review D* **48**, R2373–R2376.

Unruh, W. (1991). 'No Time and Quantum Gravity', in R. Mann & P. Wesson (eds.) *Gravitation: A Banff Summer Institute* (Singapore: World Scientific, 1991), pp. 260–75.