Physics meets philosophy at the Planck scale

Contemporary theories in quantum gravity

edited by Craig Callender

University of California at San Diego

Nick Huggett

University of Illinois at Chicago

CAMBRIDGE
UNIVERSITY PRESS

Contents

1 Introduction

Craig Callender and Nick Huggett

In recent years it has sometimes been difficult to distinguish between articles in quantum gravity journals and articles in philosophy journals. It is not uncommon for physics journals such as *Physical Review D*, *General Relativity and Gravitation* and others to contain discussion of philosophers such as Parmenides, Aristotle, Leibniz, and Reichenbach; meanwhile, *Philosophy of Science*, *British Journal for the Philosophy of Science* and others now contain papers on the emergence of spacetime, the problem of time in quantum gravity, the meaning of general covariance, etc. At various academic conferences on quantum gravity one often finds philosophers at physicists' gatherings and physicists at philosophers' gatherings. While we exaggerate a little, there is in recent years a definite trend of increased communication (even collaboration) between physicists working in quantum gravity and philosophers of science. What explains this trend?

Part of the reason for the connection between these two fields is no doubt negative: to date, there is no recognized experimental evidence of characteristically *quantum* gravitational effects. As a consequence, physicists building a theory of quantum gravity are left without direct guidance from empirical findings. In attempting to build such a theory almost from first principles it is not surprising that physicists should turn to theoretical issues overlapping those studied by philosophers.

But there is also a more positive reason for the connection between quantum gravity and philosophy: many of the issues arising in quantum gravity are genuinely philosophical in nature. Since quantum gravity forces us to challenge some of our deepest assumptions about the physical world, all the different approaches to the subject broach questions discussed by philosophers. How should we understand general relativity's general covariance – is it a significant physical principle, or is it merely a question about the language with which one writes an equation? What is the nature of time and change? Can there be a theory of the universe's boundary conditions? Must space and time be fundamental? And so on. Physicists thinking about these issues have noticed that philosophers have investigated each of them. (Philosophers have discussed the first question for roughly 20 years; the others for at least 2,500 years.) Not surprisingly, then, some physicists have turned to the work of classic and contemporary philosophers to see what they have been saying about time, space, motion, change and so on. Some philosophers, noticing this work, have responded by studying quantum gravity. They have diverse motives: some hope that their logical skills and acquaintance with such topics may serve the physicists in their quest for a theory of quantum gravity; others hope that work in the field may shed some light on these ancient questions, in the way that modern physics has greatly clarified other traditional areas of metaphysics, and still others think of quantum gravity as an intriguing 'case study' of scientific discovery in practice. In all these regards, it is interesting to note that Rovelli (1997) explicitly and positively draws a parallel between the current interaction between physics and philosophy and that which accompanied the scientific revolution, from Galileo to Newton.

This volume explores some of the areas that philosophers and physicists have in common with respect to quantum gravity. It brings together some of the leading thinkers in contemporary physics and philosophy of science to introduce and discuss philosophical issues in the foundations of quantum gravity. In the remainder of this introduction we aim to sketch an outline of the field, introducing the basic physical ideas to philosophers, and introducing philosophical background for physicists. We are especially concerned with the questions: Why should there be a quantum theory of gravity? What are the leading approaches? And what issues might constitute the overlap between quantum gravity and philosophy?

More specifically, the plan of the Introduction is as follows. Section 1.1 sets the stage for the volume by briefly considering why one might want a quantum theory of gravity in the first place. Section 1.2 is more substantive, for it tackles the question of whether the gravitational field *must* be quantized. One often hears the idea that it is actually inconsistent with known physics to have a world wherein the gravitational field exists unquantized. But is this right? Section 1.2.1 considers an interesting argument which claims that if the world exists in a half-quantized and half-unquantized form, then either superluminal signalling will be allowed or energy–momentum will not be conserved. Section 1.2.2 then takes up the idea of so-called 'semiclassical' quantum gravity. We show that the arguments for quantizing gravity are not conclusive, but that the alternative is not particularly promising either. We feel that it is important to address this issue so that readers will understand how one is led to consider the kind of theories – with their extraordinary conceptual difficulties – discussed in the book. However, those not interested in pursuing this issue immediately are invited to skip ahead to Section 1.3, which outlines (and hints at some conceptual problems with) the two main theories of quantum gravity, superstring theory and canonical quantum gravity. Finally, Section 1.4 turns to the question of what quantum gravity and philosophy have to say to each other. Here, we discuss in the context of the papers in the volume many of the issues where philosophers and physicists have interests that overlap in quantum gravity.

A word to the wise before we begin. Because this is a book concerned with the philosophical dimensions of quantum gravity, our contributors stress philosophical discussion over accounts of state-of-the-art technical developments in physics: especially, loop quantum gravity and M-theory are treated only in passing (see Rovelli 1998 and Witten 1997 for reviews, and Major 1999 for a very accesible formal introduction). Aside from sheer constraints on space, the reasons for this emphasis are two-fold: First, developments at the leading edge of the field occur very fast and do not always endure; and second, the central philosophical themes in the field can (to a large extent) be understood and motivated by consideration of the core parts of the theory that have survived subsequent developments. We have thus aimed to provide an introduction to the philosophy of quantum gravity that will retain its relevance as the field evolves: hopefully, as answers are worked out, the papers here will still raise the important questions and outline their possible solutions. But the reader should be aware that there will have been important advances in the physics that are not reflected in this volume: we invite them to learn here what issues are philosophically interesting about quantum gravity, and then discover for themselves how more recent developments in physics relate to those issues.

1.1 Why quantum gravity?

We should emphasize at the outset that currently there is no quantum theory of gravity in the sense that there is, say, a quantum theory of gauge fields. 'Quantum gravity' is merely a placeholder for whatever theory or theories eventually manage to bring together our theory of the very small, quantum mechanics, with our theory of the very large, general relativity. This absence of a theory might be thought to present something of an impediment to a book supposedly on its foundations. However, there do exist many more-or-less developed approaches to the task – especially superstring theory and canonical quantum gravity (see Section 1.3) – and the assumptions of these theories and the difficulties they share can be profitably studied from a variety of philosophical perspectives.

First, though, a few words about why we ought to expect there to be a theory of quantum gravity. Since we have no unequivocal experimental evidence conflicting with either general relativity or quantum mechanics, do we really need a quantum theory of gravitation? Why can't we just leave well enough alone, as some philosophical approaches to scientific theories seem to suggest?

It might be thought that 'instrumentalists' are able to ignore quantum gravity. Instrumentalism, as commonly understood, conceives of scientific theories merely as tools for prediction. Scientific theories, on this view, are not (or ought not to be) in the business of providing an accurate picture of reality in any deeper sense. Since there are currently no observations demanding a quantum gravitational theory, it might be thought that advocates of such a position would view the endeavour as empty and misguided speculation, perhaps of formal interest, but with no physical relevance.

However, while certain thinkers may indeed feel this way, we don't think that instrumentalists can safely ignore quantum gravity. It would be unwise for them to construe instrumentalism so narrowly as to make it unnecessary. The reason is that some of the approaches to the field may well be testable in the near future. The work that won first prize in the 1999 Gravity Research Foundation Essay Competition, for instance, sketches how both photons from distant astrophysical sources and laboratory experiments on neutral kaon decays may be sensitive to quantum gravitational effects (Ellis et al. 1999). And Kane (1997) explains how possible predictions of superstring theory – if only the theory was sufficiently tractable for them to be made – could be tested with currently available technologies. We will never observe the effects of gravitational interactions between an electron and a proton in a hydrogen atom (Feynman 1995, p. 11, calculates that such interaction would change the wave function phase by a tiny 43 arcseconds in $100T$, where T is the age of the universe!), but other effects may be directly or indirectly observable, perhaps given relatively small theoretical or experimental advances. Presumably, instrumentalists will want physics to be empirically adequate with respect to these phenomena. (We might also add the common observation that since one often doesn't know what is observable until a theory is constructed, even an instrumentalist should not restrict the scope of new theories to extant evidence.)

Another philosophical position, which we might dub the 'disunified physics' view might in this context claim that general relativity describes certain aspects of the world, quantum mechanics other distinct aspects, and that would be that. According to this view, physics (and indeed, science) need not offer a single universal theory encompassing all physical phenomena. We shall not debate the correctness of this view here, but we would like to point out that if physics aspires to provide a complete account of the world, as it traditionally has, then there must be a quantum theory of gravity. The simple reason is that general relativity and quantum mechanics cannot both be correct even in their domains of applicability.

First, general relativity and quantum mechanics cannot both be universal in scope, for the latter strictly predicts that all matter is quantum, and the former only describes the gravitational effects of classical matter: they cannot both take the whole (physical) world as their domain of applicability. But neither is the world split neatly into systems appropriately described by one and systems appropriately described by the other. For the majority of situations treated by physics, such as electrons or planets, one can indeed get by admirably using only one of these theories: for example, the gravitational effects of a hydrogen nucleus on an electron are negligible, as we noted above, and the quantum spreading of the wavepacket representing Mercury won't much affect its orbit. But in principle, the two theories govern the same systems: we cannot think of the world as divided in two, with matter fields governed by quantum mechanics evolving on a curved spacetime manifold, itself governed by general relativity. This is, of course, because general relativity, and in particular, the Einstein field equation

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.1}$$

couples the matter–energy fields in the form of the stress–energy tensor, $T_{\mu\nu}$, with the spacetime geometry, in the form of the Einstein tensor, $G_{\mu\nu}$. Quantum fields carry energy and mass; therefore, if general relativity is true, quantum fields distort the curvature of spacetime and the curvature of spacetime affects the motion of the quantum fields. If these theories are to yield a complete account of physical phenomena, there will be no way to avoid those situations – involving very high energies – in which there are non-negligible interactions between the quantum and gravitational fields; yet we do not have a theory characterizing this interaction. Indeed, the influence of gravity on the quantum realm is an experimental fact: Peters et al. (1999) measured interference between entangled systems following different paths in the Earth's gravitational field to measure gravitational acceleration to three parts in 10^9. Further, we do not know whether new low energy, non-perturbative, phenomena might result from a full treatment of the connection between quantum matter and spacetime. In general, the fact that gravity and quantum matter are inseparable 'in principle' will have *in practice* consequences, and we are forced to consider how the theories connect.

One natural reaction is to correct this 'oversight' and extend quantum methods to the gravitational interaction in the way that they were applied to describe the electromagnetic and nuclear interactions of matter, yielding the tremendously successful 'standard model' of quantum field theory. One way to develop this approach is to say that the spacetime metric, $g_{\mu\nu}$, be broken into two parts, $\eta_{\mu\nu} + h_{\mu\nu}$, representing a flat background spacetime and a gravitational disturbance respectively; and that we look for a quantum field theory of $h_{\mu\nu}$ propagating in a flat spacetime described by $\eta_{\mu\nu}$. However, in contrast to the other known forces, it turns out that all unitary local quantum field theories for gravity are non-renormalizable. That is, the coupling strength parameter has the dimensions of a negative power of mass, and so standard arguments imply that the divergences that appear in perturbative calculations of physical quantities cannot be cancelled by rescaling a finite number of physical parameters: ultimately the theory depends on an infinite number of quantities that would need to be fixed empirically. More troubling is the strong suggestion from study of the 'renormalization group' that such non-renormalizable theories become pathological at short distances (e.g. Weinberg 1983) – perhaps not too surprising a result for a theory which attempts in some sense to 'quantize distance'.

Thus the approach that worked so well for the other forces of nature does not seem applicable to gravity. Some new strategy seems in order if we are to marry quantum theory and relativity. The different programmes – both the two main ones, canonical quantum gravity and superstring theory, and alternatives such as twistor theory, the holographic hypothesis, non-commutative geometry, topological quantum field theory, etc. – all explore different avenues of attack. What goes, of course, is the picture of gravity as just another quantum field on a flat classical spacetime – again, not too surprising if one considers that there is no proper distinction between gravity and spacetime in general relativity. But what is to be expected, if gravity will not fit neatly into our standard quantum picture of the world, is that developing quantum gravity will require technical and philosophical revolutions in our conceptions of space and time.

1.2 Must the gravitational field be quantized?

1.2.1 No-go theorems?

Although a theory of quantum gravity may be unavoidable, this does not automatically mean that we must *quantize* the classical gravitational field of general relativity. A theory is clearly needed to characterize systems subject to strong quantum and gravitational effects, but it does not follow that the correct thing to do is to take classical relativistic objects such as the Riemann tensor or metric field and quantize them: that is, make them operators subject to non-vanishing commutation relations. All that follows from Section 1.1 is that a new theory is needed – nothing about the nature of this new theory was assumed. Nevertheless, there are arguments in the literature to the effect that it is inconsistent to have quantized fields interact with non-quantized fields: the world cannot be half-quantized-and-half-classical. If correct, given the (apparent) necessity of quantizing matter fields, it would follow that we must also quantize the gravitational field. We would like to comment briefly on this type of argument, for we believe that they are interesting, even if they fall short of strict no-go theorems for any half-and-half theory of quantum gravity.

We are aware of two different arguments for the necessity of quantizing fields that interact with quantum matter. One is an argument (e.g. DeWitt 1962) based on a famous paper by Bohr and Rosenfeld (1933) that analysed a semiclassical theory of the electromagnetic field in which 'quantum disturbances' spread into the classical field. These papers argue that the quantization of a given system implies the quantization of any system to which it can be coupled, since the uncertainty relations of the quantized field 'infect' the coupled non-quantized field. Thus, since quantum matter fields interact with the gravitational field, these arguments, if correct, would prove that the gravitational field must also be quantized. We will not discuss this argument here, since Brown and Redhead (1981) contains a sound critique of the 'disturbance' view of the uncertainty principle underpinning these arguments.

Interestingly, Rosenfeld (1963) actually denied that the 1933 paper showed any inconsistency in semiclassical approaches. He felt that empirical evidence, not logic, forced us to quantize fields; in the absence of such evidence 'this temptation [to quantize] must be resisted' (1963, p. 354). Emphasizing this point, Rosenfeld ends his paper with the remark, 'Even the legendary Chicago machine cannot deliver sausages if it is not supplied with hogs' (1963, p. 356). This encapsulates the point of view we would like to defend here.

The second argument, which we will consider, is due to Eppley and Hannah (1977) (but see also Page and Geilker 1981 and Unruh 1984). The argument – modified in places by us – goes like this. Suppose that the gravitational field were relativistic (Lorentzian) and classical: not quantized, not subject to uncertainty relations, and not allowing gravitational states to superpose in a way that makes the classical field indeterminate. The contrast is exactly like that between a classical and quantum particle.[1] Let us also momentarily assume the standard interpretation of quantum mechanics, whereby a measurement interaction instantaneously collapses the wave function into an eigenstate of the relevant observable. (See, for example Aharonov and Albert 1981, for a discussion of the plausibility of this interpretation in the relativistic context.)

Now we ask how this classical field interacts with quantized matter, for the moment keeping all possibilities on the table. Eppley and Hannah (1977) see two (supposedly) exhaustive cases: gravitational interactions either collapse or do not collapse quantum states.

Take the first horn of the dilemma: suppose the gravitational field *does not* collapse the quantum state of a piece of matter with which it interacts. Then we can send superluminal signals, in violation of relativity, as conventionally understood. Eppley and Hannah (1977) (and Pearle and Squires 1996) suggest some simple ways in which this can be accomplished using a pair of entangled particles, but we will use a modification of Einstein's 'electron in a box' thought experiment. However, the key to these examples is the (seemingly unavoidable) claim that if a gravitational interaction does not collapse a quantum state, then the dynamics of the interaction depend on the state. In particular, the way a classical gravitational wave scatters off a quantum object would depend on the spatial wave function of the object, much as it would depend on a classical mass distribution. Thus, scattering experiments are at least sensitive to changes in the wave function, and at best will allow one to determine the form of the wave function – without collapsing it. It is not hard to see how this postulate, together with the usual interpretation of quantum measurements, allows superluminal signalling.

We start with a rectangular box containing a single electron (or perhaps a microscopic black hole), in a quantum state that makes it equally likely that the electron will be found in either half of the box. We then introduce a barrier between the two halves and separate them, leaving the electron in a superposition of states corresponding to being in the left box and being in the right box. If the probabilities of being in each box are equal, then the state of the particle will be:

$$\psi(x) = 1/\sqrt{2}(\psi_L(x) + \psi_R(x)), \tag{1.2}$$

where $\psi_L(x)$ and $\psi_R(x)$ are wave functions of identical shape but with supports inside the left and right boxes respectively.

Next we give the boxes to two friends Lefty and Righty, who carry them far apart (without ever looking in them of course). In Einstein's original version (in a letter discussed by Fine 1986, p. 35–39, which is a clarification of the EPR argument in its published form), when Lefty looked inside her box – and say found it empty – an element of reality was instantaneously present in Righty's box – the presence of the electron – even though the boxes were spacelike separated. Assuming the collapse postulate, when Lefty looks in her box a state transition,

$$1/\sqrt{2}(\psi_L(x) + \psi_R(x)) \rightarrow \psi_R(x) \tag{1.3}$$

occurs. In the familiar way, either some kind of spooky non-local 'action' occurs or the electron was always in Righty's box and quantum mechanics is incomplete, since $\psi(x)$ is indeterminate between the boxes. Of course, this experiment does not allow signalling, for if Righty now looks in his box and sees the electron, he could just as well conclude that he was the first to look in the box, collapsing the superposition. And the long run statistics generated by repeated measurements that Righty observes will be 50 : 50, electron : empty, whatever Lefty does – they can only determine the correlation by examining the joint probability distribution, to which Righty, at his wing, does not have access.

In the present case the situation is far more dire, for Righty can use our non-collapsing gravitational field to 'see' what the wave function in his box is without collapsing it. We simply imagine that the right-hand box is equipped with apertures that allow gravitational waves in and out, and that Righty arranges a gravitational wave source at one of them and detectors at the others.[2]

Since the scattering depends on the form of the wave function in the box, any changes in the wave function will show up as changes in the scattering pattern registered by the detectors. Hence, when Lefty now looks in her box – and suppose this time she finds the electron – Righty's apparatus will register the collapse instantaneously; there will be no scattering source at all, and the waves will pass straight through Righty's box. That is, before Lefty looks, the electron wave function is $\psi(x)$ and Righty's gravity wave scatters off $\psi_R(x)$; after Lefty collapses the electron, its state is $\psi_L(x) + 0$ and so Righty's gravity wave has no scattering source. And since we make the usual assumption that the collapse is instantaneous, the effect of looking in the left box is registered on the right box superluminally. So, if Righty and Lefty have a prior agreement that if Lefty performs the measurement then she fancies a drink after work, otherwise she wants to go to the movies, then the apparatus provides Righty with information about Lefty's intentions at a spacelike separated location.[3]

It is crucial to understand that this experiment is *not* a variant of 'Wigner's friend'. One should absolutely not think that scattering the gravity wave off the electron wave function leads to an entangled state in which the gravity wave is in a quantum superposition, which is itself collapsed when measured by the detectors, producing a consequent collapse in the electron wave function. Of course, such things might occur in a theory of quantum gravity, but they cannot occur in the kind of theory that we are presently discussing: a theory with a *classical* gravitational field, which just means a theory in which there are no quantum superpositions of the gravitational field. There is in this theory no way of avoiding signalling by introducing quantum collapses of the gravitational field, since there is nothing to collapse.

It is also important to see how the argument depends on the interpretation of quantum mechanics. On the one hand it does not strictly require the standard interpretation of quantum mechanics, but can be made somewhat more general. In our example, the component of the wave function with support on Righty's box went from $\psi_R(x)$ to 0, which is a very sharp change. But the argument doesn't need a sharp change, it just needs a detectable change, to $\epsilon\psi_R(x)$, say. On the other hand, it is necessary for the argument that normal measurements can produce effects at spacelike separated regions. For then the gravitational waves provide an abnormal way of watching a wave function without collapsing it, to see when such effects occur. Thus, an interpretation of quantum mechanics that admits a dynamics which prevents superluminal propagation of any disturbance in the wave function will escape this argument. Any no-collapse theory whose wave function is governed at all times by a relativistic wave equation will be of this type.

The conclusion of this horn of the dilemma is then the following. If one adopts the standard interpretation of quantum mechanics, and one claims that the world is divided into classical (gravitational) and quantum (matter) parts, and one models quantum–classical interactions without collapse, then one must accept the possibility of superluminal signalling. And further, though practical difficulties may prevent one from ever building a useful signalling device, the usual understanding of relativity prohibits superluminal signalling, even in principle. Of course, this interpretation of relativity is a subtle matter in a number of ways, for instance concerning the possibility of Lorentz-invariant signalling (Maudlin 1994) and even the possibility of time travel (see, e.g. Earman 1995a). And of course, given the practical difficulty of performing such an experiment, we do not have definitive empirical grounds for ruling out such signalling. But since the kind of signalling described here could pick out a preferred foliation of spacetime – on which the collapse occurs – it does violate relativity in an important sense. Thus, someone who advocates a standard interpretation of quantum mechanics, a half-and-half view of the world and a no-collapse theory of classical–quantum interactions must deny relativity as commonly understood. They would need a very different theory that could accommodate the kind of superluminal signalling demonstrated, but that also approximates the causal structure of general relativity in all the extant experiments. (Note that this conclusion is in line with our earlier, more general, argument for the existence of a theory of quantum gravity; and note that the present argument really only demonstrates the need for a new theory – it does not show that quantizing the metric field is the only way to escape this problem.)

Of course, as mentioned, one *might* be able to avoid this horn of the dilemma by opting for a no-collapse interpretation of quantum mechanics, e.g. some version of Bohmian mechanics, or Everettian theories. We are not aware of any actual proposal for a half-and-half world that exploits this possibility (e.g. Bohmian quantum gravity – see below – aims to quantize the gravitational field). But the space may exist in the logical geography. In Bohm's theory, however measurements can have non-local effects on particle positions. Signalling could therefore occur if scattering at the gravitational field depended on the particle configuration and not only the wave function.

Let's turn to the other horn of the dilemma, where now we suppose that gravitational interactions can collapse quantum states of matter. Interestingly enough, there are a number of concrete suggestions that gravity should be thus implicated in the measurement problem, so it is perhaps not too surprising that attempts to close off this horn are, if anything, even less secure.

Eppley and Hannah's (1977) argument against a collapsing half-and-half theory is that it entails a violation of energy–momentum conservation. First, we assume that when our classical gravitational wave scatters off a quantum particle its wave function collapses, to a narrow Gaussian say. Second, we assume that the gravity wave scatters off the collapsed wave function as if there were a point particle localized at the collapse site. Then the argument is straightforward: take a quantum particle with sharp momentum but uncertain position, and scatter a gravity wave off it. The wave function collapses, producing a localized particle (whose position is determined by observing the scattered wave), but with uncertain momentum according to the uncertainty relations. Making the initial particle slow and measuring the scattered gravity wave with sufficient accuracy, one can pinpoint the final location sharply enough to ensure that the uncertainty in final momentum is far greater than the sharp value of the initial momentum. Eppley and Hannah conclude that we have a case of momentum non-conservation, at least on the grounds that a subsequent momentum measurement could lead to a far greater value than the initial momentum. (Or perhaps, if we envision performing the experiment on an ensemble of such particles, we have no reason to think that the momentum expectation value after will be the same as before.)

As with the first argument, the first thing that strikes one about this second argument is that it does not obviously depend on the fact that it is an interaction with the gravitational field that produces collapse. Identical reasoning could be applied to any sufficiently high resolution particle detector, given the standard collapse interpretation of measurement. Since this problem for the collapse interpretation is rather obvious, we should ask whether it has any standard response. It seems that it does: as long as the momentum associated with the measuring device is much greater than the uncertainty it produces, then we can sweep the problem under the rug. The non-conservation is just not relevant to the measurement undertaken. If this response works for generic measurements, then we can apply it in particular to gravitationally induced collapse, leaving Eppley and Hannah's argument inconclusive.

But how satisfactory is this response in the generic case? Just as satisfactory as the basic collapse interpretation: not terribly, we would say. Without rehearsing the familiar arguments, 'sweeping quantities under the rug' in this way seems troublingly *ad hoc*, pointing to some missing piece of the quantum puzzle: hidden variables perhaps or, as we shall consider here, a precise theory of collapse. Without some such addition to quantum mechanics it is hard to evaluate whether such momentum non-conservation should be taken seriously or not, but with a more detailed collapse theory it is possible to pose some determinate questions. Take, as an important example, the 'spontaneous localization' approaches of Ghirardi, Rimini, and Weber (1986) or, more particularly here, of Pearle and Squires (1996). In their models, energy is indeed not conserved in collapse, but with suitable tuning (essentially smearing matter over a fundamental scale), the effect can be made to shrink below anything that might have been detected to date.[4]

Whether such an answer to non-conservation is satisfactory depends on whether we must take the postulate of momentum conservation as a fundamental or experimental fact, which in turn depends on our reasons for holding the postulate. In quantum mechanics, the reasons are of course that the spacetime symmetries imply that the self-adjoint generators of temporal and spatial translations commute, $[\hat{H}, \hat{P}] = 0$, and the considerations that lead us to identify the generator of spatial transformations with momentum (cf., e.g. Jordan 1969). The conservation law, $d\langle\hat{P}\rangle/dt = 0$ then follows simply. But of course, implicit in the assumption that there is a self-adjoint generator for temporal translations, $\hat{H}$, is the assumption that the evolution operator, $\hat{U}(t) = e^{-i\hat{H}t/\hbar}$, is unitary. But in a collapse, it is exactly this assumption that breaks down: so what Eppley and Hannah in fact show is only that in a collapse our fundamental reasons for expecting momentum conservation fail. But if all that remains are our empirical reasons, then the spontaneous localization approaches are satisfactory on this issue, as are other collapse models that hide momentum non-conservation below the limits of observation. Thus, the incompleteness problem aside, sweeping momentum uncertainty under the rug need not do any harm.[5] In this respect, it is worth noting that if gravitational waves cause quantum jumps, then the effect must depend in some way on the strength of the waves. The evidence for this assertion is the terrestrial success of quantum mechanics despite the constant presence on Earth of gravity waves from deep space sources (and indeed from the motions of local objects). If collapse into states sharp in position were occurring at a significant rate then, for instance, we would not expect matter to be stable, since energy eigenstates are typically not sharp in position, nor would we observe electron diffraction, since electrons with sharp positions move as localized particles, without interfering.[6]

Since the size of the collapse effect must depend on the gravitational field in some way, one could look for a theory in which the momentum of a gravitational wave was always much larger than the uncertainty in any collapse it causes. Then any momentum non-conserving effect would be undetectable, and Eppley and Hannah's qualms could again be swept under the rug.

Indeed, the spontaneous localization model developed by Pearle and Squires (1996) has these features to some degree. They do not explicitly model a collapse caused by a gravitational wave, but rather use a gravitational field whose source is a collection of point sources, which punctuate independently in and out of existence. In their model, the rate at which collapse occurs depends (directly) on the mass of the sources and (as a square root) on the probability for source creation at any time, so that a stronger source field produces a stronger gravitational field and a greater rate of collapse. And the amount of energy produced by collapse is undetectable by present instruments, given a suitable fixing of constants, sweeping it under the rug. (Pearle and Squires also argue that the collapse rate is great enough to prevent signalling.)

Finally, Roger Penrose also links gravity to collapse (see Chapter 13). He in fact advocates a model in which gravity is quantized, but this is not crucial to the measure of collapse rate he offers. He proposes that the time rate of collapse of a superposition of two separated wave packets, T, is determined by the (Newtonian) gravitational self-energy of the difference of the two packets, E_Δ: $T \sim \hbar/E_\Delta$. Penrose also proposes a test for this model, which Joy Christian criticizes and refines in Chapter 14.

As we said at the start of this section, arguments in the style of Eppley and Hannah do not constitute no-go theorems against half-and-half theories of quantum gravity. There are ways to evade both horns of the dilemma: adopting a no-collapse theory could preclude superluminal signalling in the first horn, and allowing for unobservable momentum non-conservation makes the result of the second horn something one could also live with. This particular argument, which is often repeated in the literature and in conversation, fails. Though we have not shown it, we would like to here register our skepticism that any argument in the style of Eppley and Hannah's could prove that it is inconsistent to have a world that is part quantum and part classical. Rosenfeld (1963) is right. Empirical considerations must create the necessity, if there is any, of quantizing the gravitational field.

Even so, the mere possibility of half-and-half theories does not make them attractive, and aside from their serious attention to the measurement problem, it is important to emphasize that they have not yielded the kind of powerful new insights that attract large research communities. Note too that most physicists arguing for the necessity of quantum gravity do not take the above argument as the main reason for quantizing the gravitational field. Rather, they usually point to a list of what one might call methodological points in favour of quantum gravity; see Chapter 2 for one such list. These points typically include various perceived weaknesses in contemporary theory, and find these sufficiently suggestive of the need for a theory wherein gravity is quantized. We have no qualms with this kind of argument, so long as it is recognized that the need for such a theory is not one of logical or (yet) empirical necessity.

1.2.2 The semiclassical theory

Finally, we should discuss a specific suggestion for a half-and-half theory due to Møller (1962) and Rosenfeld (1963), which appears – often as a foil – in the literature from time to time (see Chapters 2 and 13). This theory, 'semiclassical quantum gravity' (though any half-and-half theory is in some sense semiclassical), postulates first that the spacetime geometry couples to the expectation value of the stress–energy tensor:

$$G_{\mu\nu} = 8\pi \langle \hat{T}_{\mu\nu} \rangle_{\Psi}. \tag{1.4}$$

$G_{\mu\nu}$ is the *classical* Einstein tensor and $\langle \hat{T}_{\mu\nu} \rangle_{\Psi}$ is the expectation value for the stress–energy operator given that the *quantum* state of the matter fields is Ψ. Clearly this is the most obvious equation to write down given the Einstein field equation of classical general relativity, and given quantum rather than classical matter: $\langle T_{\mu\nu} \rangle_{\Psi}$ is the most obvious 'classical' quantity that can be coupled to $G_{\mu\nu}$.

Now, eqn. 1.4 differs importantly from the classical equation (eqn. 1.1), in that the latter is supposed to be 'complete', coding all matter–space and matter–matter dynamics: in principle, no other dynamical equations are required. Equation 1.4 cannot be complete in this sense, for it only imposes a relation between the spacetime geometry and the *expectation value* of the matter density, but typically many quantum states share any given expectation value. For example, being given the energy expectation value of some system as a function of time does not, by itself, determine the evolution of the quantum state (one also needs the standard connection between the Hamiltonian and the dynamics). So the semiclassical theory requires a separate specification of the quantum evolution of the matter fields: a Schrödinger equation on a curved spacetime. But this means that semiclassical quantum gravity is governed by an unpleasantly complicated dynamics: one for which we must seek 'self-consistent' solutions to two disparate of motions. Finding a model typically proceeds by first picking a spacetime – say a Schwarzschild black hole – and solving the Schrödinger equation for the matter fields on the spacetime. But this ignores the effect of the field on the spacetime, so next one wants to find the stress–energy tensor for this solution, and plug that into the semiclassical equation, to find corrections to the original spacetime. But then the assumption of a Schwarzschild solution no longer holds, and the Schrödinger equation must be solved for the new geometry, giving a new stress–energy expectation value to be fed back into the semiclassical equation, and so on and so on. What one of course hopes is that this process converges on a spacetime and matter field that satisfy both equations, but in the absence of such solutions it is not even clear that the equations are mutually consistent.

This lack of unity is one reason that physicists by-and-large do not take the semiclassical theory seriously as a 'fundamental' theory. The idea instead is that it can be used as an heuristic guide to some suggestive results in the absence of a real theory of quantum gravity; in the famous case mentioned, 'Hawking radiation' is produced by quantum fields in a Schwarzschild spacetime, and if one could feed the 'back reaction' into the semiclassical equation one would expect to find the black hole radius decreasing as it evaporated. Unfortunately, a number of technical problems face even this example, and though research is active and understanding is increasing, there are still no known interesting solutions with evolving spacetimes in four dimensions (see Wald 1994 for a discussion).

Things look even worse when one considers the collapse dynamics for quantum mechanics. On the standard interpretation, the unitary dynamics for a system must be supplemented with a collapse during certain 'measurement' interactions. If the semiclassical theory were complete, then not only would the unitary evolution of quantum matter need to be contained in eqn. 1.4, but so would the collapse. But – turning around an argument of Unruh's (1984) – it is impossible for eqn. 1.4 to contain a sharp collapse: the Einstein tensor is necessarily conserved, $G_{\mu;\nu}{}^{\nu} = 0$, but a collapse would lead to a discontinuity in the stress–energy expectation value, $\langle \hat{T}_{\mu}{}^{\nu} \rangle_{;\nu} \neq 0$. Nor does it seem plausible that eqn. 1.4 contains a smoother collapse: why should measuring events invariably produce the appropriate change in $G_{\mu\nu}$? In that case, the theory needs to be supplemented with a collapse dynamics, perhaps along the lines suggested by Pearle and Squires, but certainly constrained by the arguments considered in Section 1.2.1. In this context one point is worth noting: if the collapse mechanism is sharp then Unruh's argument shows either that eqn. 1.4 is incorrect, or that the LHS is not a tensor field everywhere, but only on local patches of spacetime, with 'jumps' in the field outside the patches.[7]

Our assessment of the semiclassical theory *as a candidate fundamental theory* is much the same as for half-and-half theories in general: while they are not impossible, if one weighs the insights they offer against the epicycles they require for their maintenance, then they do not appear to be terribly progressive. Certainly they take seriously the measurement problem, and so address arguments by, for instance, Penrose (1989) that gravity and collapse are interrelated. And certainly, the semiclassical theory has provided an invaluable and revealing tool for exploring the boundary between general relativity and quantum mechanics, yielding a picture of what phenomena – such as black hole thermodynamics – might be expected of a theory of quantum gravity. But on the other hand, arguing that half-and-half theories are fundamental involves more negotiating pitfalls than producing positive results. Thus it is not surprising that, although half-and-half theories have not been shown to be inconsistent, they are not the focus of most work in quantum gravity.

1.3 Approaches to quantum gravity

As we mentioned earlier, there currently is no quantum theory of gravity. There are, however, some more-or-less developed approaches to the field, of which the most actively researched fall into one of two broad classes, superstring theory and canonical quantum gravity. Correspondingly, most of the chapters in this book that deal explicitly with current research in the field discuss one or the other of these programmes; thus it will be useful to give here preparatory sketches of both classes.

(A warning: more space is devoted here to the canonical approach than to string theory. This does not mirror their relative popularities in the contemporary physics community, but reflects the greater development of philosophical discussion within the canonical programme.)

1.3.1 Superstrings

First, superstring theory, which is discussed in Chapters 5, 6, and 7. Superstring theory seeks to provide a unified quantum theory of all interactions, in which the elementary entities are one-dimensional extended objects (strings), not point particles. Superstring theory arose from work on the strong interaction in the late 1960s and early 1970s when it was shown that all the properties of a certain interesting model of the strong interaction (Veneziano's model) could be duplicated by a Lagrangian theory of a relativistic string. Though interesting, this idea did not really take off. In the mid-1970s, however, it was demonstrated that graviton–graviton (quanta of the gravitational field) scattering amplitudes were the same as the amplitudes of a certain type of closed string. This fact led to the idea that superstring theory is a theory of all forces, not only the strong interaction.

Superstrings, consequently, are not meant to represent single particles or single interactions, but rather they represent the entire spectrum of particles through their vibrations. In this way, superstring theory promises a novel and attractive 'ontological unification'. That is, unlike electrons, protons and neutrons – which together compose atoms – and unlike quarks and gluons – which together compose protons – strings would not be merely the smallest object in the universe, one from which other types of matter are composed. Strings would not be merely *constituents* of electrons or protons in the same manner as these entities are constituents of atoms. Rather, strings would be all that there is: electrons, quarks, and so on would simply be different vibrational modes of a string. In the mid-1980s this idea was taken up by a number of researchers, who developed a unitary quantum theory of gravity in ten dimensions.

A sketch of the basic idea is as follows (see Chapters 5 and 6 for more extensive treatments). Consider a classical one-dimensional string propagating in a relativistic spacetime, sweeping out a two-dimensional worldsheet, which we can treat as a manifold with 'internal' co-ordinates (σ, τ). One can give a classical treatment of this system in which the canonical variables are $X^\mu(\sigma, \tau)$, where X^μ $(\mu = 0, 1, 2, 3)$ are the spacetime co-ordinates of points on the string worldsheet. When the theory is quantized, this embedding function is treated as a quantum field theory of excitations *on the string*, $\hat{X}^\mu(\sigma, \tau)$. A number of fascinating and suggestive properties arise as necessary consequences of such quantization.

- For instance, it was found that every consistent interacting quantized superstring theory necessarily includes gravity. That is, closed strings all have gravitons (massless spin-2 particles) in their excitation spectrum and open strings contain them as intermediate states.
- In addition, the classical spacetime metric on the background spacetime must satisfy a version of Einstein's field equations (plus small perturbations) if we (plausibly) demand that $X^\mu(\sigma, \tau)$ be conformally invariant on the string. Thus, arguably, general relativity follows from string theory in the appropriate limit.

- It is also necessary that the theory be supersymmetric (a symmetry allowing transformations between bosons and fermions), a property independently attractive to physicists seeking unified theories.

So part of the motivation for string theory comes from the feeling that it is almost too good to be a coincidence that the mere requirement of quantizing a classical string automatically brings with it gravity and supersymmetry. Of course, another notorious consequence of quantizing a string is that spacetime must have a dimension n, where $n > 4$ ($n = 26$ for the bosonic string, $n = 10$ when fermions are added). But even here it can be claimed that the extra spacetime dimensions do not arise by being put in artificially (as is the case, arguably, in Kaluza–Klein theory). Rather, they again arise as a necessary condition for consistent quantization.

More recently, since the mid-1990s, string theorists have explored various symmetries known as 'dualities' in order to find clues to the non-perturbative aspects of the theory. These dualities, they believe, hint that the five different existing classes of string theories may in fact just be aspects of an underlying theory sometimes called 'M-theory', where 'M' = 'magic', 'mystery', 'membrane', or we might add, 'maybe'. (In a sense, superstring theory can be seen as returning to its early foundations, since the ideas of Veneziano were based upon a [different] kind of 'duality' in his model of the strong interaction.) There have been many exciting results along these lines during the 1990s. These are briefly described in Chapters 2 and 5 (see also Witten 1997). In addition, important and impressive results concerning the thermodynamics of black holes have also been derived from the perspective of string theory, and these are described and discussed in Chapter 7.

It is perhaps legitimate to view the difference between the perturbative superstring theory of the mid-1980s and the non-perturbative M-theory of 1994–present as the difference between whether superstring theory is a new fundamental theory or not. The older superstring theory is, in a sense, not fundamental; quantum mechanics still was. Superstring theory was 'just' quantum mechanics applied to classical strings. (Of course there was no 'just' about it as regards the mathematical and physical insight needed to devise the theory!) But with today's string field theory, we see intimations of a wholly new fundamental theory in the various novel dualities and non-perturbative results. However, it is still troubling that so much of the success of string theory derives from the enormous mathematical power and elegance of the theory, rather than from empirical input. In his article, Weingard draws attention to this point, arguing that, unlike other theories which turned out to be successful, such as general relativity, string theory is not based on any obvious physical 'clues'. This article was written before the recent developments in M-theory, so we leave it to the reader to consider whether the situation has changed.

Another issue of philosophical significance, discussed in Chapters 5 and 6, concerns the nature of spacetime according to string theory. The original formulation of string theory was envisaged as an extension of perturbative quantum field theory from point particles to strings; thus, as we described earlier, strings were taken to carry the gravitational field on a flat background spacetime. (Part of the promise of this approach was that renormalization difficulties are at least rendered more tractable, and at best do not occur at all.) On this view, then, spacetime appears very much as it did classically: we have matter and forces evolving on an 'absolute', non-dynamical manifold. One disadvantage of this approach is easily overcome: as Witten describes, it is simple to generalize from the assumption of a flat background by inserting your metric of choice into the Lagrangian for the string field. This observation, plus the fact that conformal invariance for the field on the string demands that the Einstein equation be satisfied by the spacetime metric, leads Witten to propose that we should not see spacetime as an absolute background in string theory after all. Since spacetime is captured by the field theory on the string, 'one does not have to have spacetime any more, except to the extent that one can extract it from a two-dimensional field theory'. (In Chapter 6, Weingard makes a similar point in his discussion of the second quantization of string theory.) To support his claim Witten also describes a duality symmetry of string theory that identifies small circles with larger circles. The idea is that no circle can be shrunk beneath a certain scale, and so there is a minimum – quantum – size in spacetime, and so no absolute background continuum. More on this claim in the final section.

1.3.2 Canonical quantum gravity

In contrast to superstring theory, canonical quantum gravity seeks a non-perturbative quantum theory of only the gravitational field. It aims for consistency between quantum mechanics and gravity, not unification of all the different fields. The main idea is to apply standard quantization procedures to the general theory of relativity. To apply these procedures, it is necessary to cast general relativity into canonical (Hamiltonian) form, and then quantize in the usual way. This was (partially) successfully done by Dirac (1964) and (differently) by Arnowitt, Deser, and Misner (1962). Since it puts relativity into a more familiar form, it makes an otherwise daunting task seem hard but manageable. In the remainder of this section we will give an intuitive sketch of the steps involved in this process, but be aware that many (unsolved) difficulties lie in the way of its successful completion. The reader should also be warned that we introduce these ideas with an out-of-date formulation of the theory, namely, the geometrodynamical formulation. More sophisticated and successful formulations exist – notably, the Ashtekar variable and loop variable approaches – but as these do not by themselves significantly affect the philosophical issues facing canonical quantum gravity, we here confine ourselves to the simpler and more intuitive picture.

In the standard Hamiltonian formulation (all this material is covered more fully in Chapter 10), one starts with canonical variables – say the position, x, and momentum, p, of a particle – which define the appropriate phase space for the system. Given a point in the phase space – say the instantaneous position and momentum of the particle – the Hamiltonian, $H(x, p)$, for the system will generate a unique trajectory with respect to a time parameter. So, to apply this approach to general relativity the first job is to define the relevant variables. The intuitive picture sought is one in which a three-dimensional spatial manifold, Σ, evolves through an arbitrary time parameter τ, so the natural thing is to decompose spacetime into space and time. In the geometrodynamical formulation, a spatial 3-metric $h_{ab}(x)$ on Σ plays the role of the canonical position, and a canonically conjugate momentum $p^{ab}(x)$ is also defined (it is closely related to the extrinsic curvature of Σ in the spacetime). The phase space for this system is thus the space of all possible 3-spaces and conjugate momenta, so the pair (h, p) fix an instantaneous state of Σ. (It is worth noting that the foliation of spacetime implied by this procedure is at odds with the central tenets of general relativity and will be a source of difficulties further down the line.) Finally, one gives a Hamiltonian that generates trajectories through phase space in agreement with relativity: trajectories such that the stack of 3-spaces form a model of Einstein's field equation.

Canonical quantization of such a system then means (in the first place) finding an operator representation of the canonical variables obeying the canonical commutation relations, say, $[\hat{x}, \hat{p}] = i\hbar$, or in our case $[\hat{h}, \hat{p}] = i\hbar$. Usually one hopes that smooth (wave) functions, $\psi(x)$, on the canonical position space (configuration space) will carry such a representation (since $\hat{x} = x$ and $\hat{p} = i\hbar\,\partial/\partial x$ are operators on such functions satisfying the commutation relations). One finally obtains a quantum Hamiltonian operator by replacing the canonical variables in the classical Hamiltonian with their operator representations: $H(x, p) \to \hat{H}(\hat{x}, \hat{p})$. States of the quantum system are of course represented by the wave functions, and evolutions are generated by the quantized Hamiltonian via the Schrödinger equation: $\hat{H}\psi = i\hbar\,\partial\psi/\partial t$. If all this went through for general relativity, then we would expect states of quantum gravity to be wave functions $\psi(h)$ over the configuration space of possible – Riemannian – 3-metrics, *Riem*(3), and physical quantities including the canonical variables to be represented as operators on this space.

The general relativistic Hamiltonian generates evolutions in the direction of the time that parameterizes the stack of 3-spaces, but there is no reason why this should be normal to any given point on a 3-space. So, if T is a vector (field) in the direction of increasing stack time in the spacetime, and n is a unit vector (field) everywhere normal to the 3-spaces, we can decompose T into normal and tangential components, $T = Nn + \vec{N}$. N is the 'lapse' function, coding the normal component of T, and $\vec{N}$ is the 'shift' vector (field), which is always tangent to the 3-space. Not surprisingly, when one works through the problem (writing down a Lagrangian, then using Hamilton's equations), the Hamiltonian for the system can be split into parts generating evolutions normal and tangent to the 3-spaces. Added together, these parts generate transformations in the direction of the time parameter. Thus, a standard Hamiltonian takes the form (for certain functions, C^{μ}, of the canonical variables):

$$H = \int d^3x\,(C \cdot N + N^i C_i). \tag{1.5}$$

Variation of h and p yields six of the Einstein equation's ten equations of motion, and variation of the N and $\vec{N}$ produces the so-called 'Hamiltonian' and 'momentum' constraints' (which hold at each point in spacetime):

$$C = C_i = 0. \tag{1.6}$$

In other words, not every point of our phase space actually corresponds to a (hypersurface of a) solution of general relativity, but only those for which (h, p)

satisfy eqn. 1.6: in the formulation general relativity is a *constrained Hamiltonian system.* As a consequence of eqn. 1.6, $H = 0$, though in the classical case at least, this does not mean that there is no dynamics: the first six equations of motion ensure that the 3-space geometry varies with time. It does however lead to some deep problems in the theory, both in the classical and quantum contexts.

The idea of a constraint is of course fairly straightforward: consider, for example, a free particle moving on a plane with freely specifiable values of position, x and y, and of momentum, p_x and p_y: there are four degrees of freedom. However, if confined to a circle, $x^2 + y^2 = a$, the particle must satisfy the constraint $xp_x + yp_y = 0$, which allows us to solve for one of the variables in terms of the others: the constrained system has only three degrees of freedom. Pictorially, the unconstrained particle's state may be represented anywhere in the four-dimensional phase space spanned by the two position and two momenta axes, but the constrained particle can only 'live' on the three-dimensional subspace – the 'constraint hypersurface' – on which the constraint holds. In this model the simplest way to approach the motion is to reparameterize the system to three variables in which the constraint is automatically satisfied: effectively making the constraint hypersurface the phase space.

Now, finding a constraint for a Hamiltonian system often (though not in the previous toy example) indicates that we are dealing with a gauge theory: there is some symmetry transformation between states in the phase space that leaves all dynamical parameters unchanged. The usual understanding is that since any physical quantities must 'make a difference' dynamically, all observables (physically real quantities) must be gauge invariant. (Note that this is a much stronger notion than a covariant symmetry, the idea that transformed quantities, though distinguishable, obey the same equations of motion.) Such systems are of course fundamental to contemporary field theory, since imposing local gauge invariance on a field requires introducing a 'connection', A, which allows comparisons of values of the field at infinitesimally separated points (working as the affine connection to allow differentiation of fields over spacetime). A acts as a second field in the gauge invariant field equation, mediating interactions; in quantum field theory, the original field represents 'matter' and the connection field represents exchange particles, such as photons (see Redhead 1983). Note however that A is an example of a gauge *non*-invariant quantity, so although it is crucial for understanding interactions, it contains unphysical degrees of freedom. This apparent paradox shows the subtleties involved in understanding gauge theories.[8]

We can apply these lessons to the Hamiltonian formulation of general relativity. The constraint appears to be connected to a symmetry, this time the general covariance of the theory, understood as diffeomorphism invariance. That is, if $\langle M, g, T\rangle$ represents a spacetime of the theory (where M is a manifold, g represents the metric field and T the matter fields), and $D*$ is a smooth invertible mapping on M, then $\langle M, D * g, D * T\rangle$ represents *the very same* spacetime. Crudely, smooth differences in how the fields are arranged over the manifold are not physically significant.

It is vital to note that we have already reached the point at which controversial philosophical stances must be taken. As Belot and Earman explain, to understand diffeomorphism this way, as a gauge symmetry, is to take a stance on Einstein's infamous 'hole argument', and hence on various issues concerning the nature of spacetime; in turn, these issues will bear on the 'problem of time', introduced below. We will maintain the gauge understanding at this point, since it is fairly conventional among physicists (as several articles here testify).

Though the matter is subtle, if intuitively plausible, the momentum and Hamiltonian constraints (eqn. 1.6) are believed to capture the invariance of general relativity under spacelike and timelike diffeomorphisms respectively (e.g. Unruh and Wald 1989). As it happens (unlike the toy case) the constraints cause the Hamiltonian to vanish. This is not atypical of generally covariant Hamiltonian systems (see Belot and Earman's discussion of a parametrized free particle in Chapter 10 for an example), but what is atypical in this theory is that the Hamiltonian is *entirely* composed of constraints.

In fact, since things have been chosen nicely so that the momentum and Hamiltonian constraints are associated with diffeomorphisms tangent and normal to the 3-space, Σ, respectively, satisfying the former with a reparameterization is easy: instead of counting every h on Σ as a distinct possibility, we take only equivalence classes of 3-spaces related by diffeomorphisms to represent distinct states. This move turns our earlier configuration space into 'superspace', so we now want quantum states to be wave functions over superspace.

Once this move is made, all that is left of the Hamiltonian is the Hamiltonian constraint. But now a reparameterization seems out of the question, for the Hamiltonian constraint is related to diffeomorphisms in the time direction. If we try to form a state space in which all states related by temporal diffeomorphisms are counted as the same, then we have no choice but to treat whole spacetimes, not just 3-spaces, as states. Otherwise it just makes no sense to pose the question of whether the two states are related by the diffeomorphism. But in this case the state space consists, not of 3-geometries, but of full solutions of general relativity. And in this case there are no trajectories of evolving solutions, and the Hamiltonian and Schrödinger pictures no longer apply.

Despite this difficulty, it is still possible to quantize our system by (more or less) following Dirac's quantization scheme and requiring that the constraint equations be satisfied as operator equations, heuristically writing:

$$\hat{C}\Psi = \hat{C}_i\Psi = 0. \tag{1.7}$$

Since the momentum constraints are automatically satisfied in superspace, the focus in canonical quantum gravity is on

$$\hat{C}\Psi = 0, \tag{1.8}$$

the so-called *Wheeler–DeWitt equation*. This equation's interpretation has generated much controversy, as we shall describe in the next section. For now, be aware that very significant conceptual and formal difficulties confront the intuitive picture we have sketched: not least that there are no known solutions to the problem as constructed so far![9] There are, however, some solutions to this equation when one writes the theory in terms of Ashtekar's so-called 'new variables'. These variables have overcome many technical obstacles to the older canonical theory, and have greatly reinvigorated the canonical programme in the past decade (see Rovelli 1998b for a review).

Finally, and extremely speculatively, we note that both superstring theory and the canonical programme have evolved greatly from their initial formulations, and, as far as we are aware, it is possible that they are converging in some way: for example, perhaps some descendent of the Ashtekar formulation will turn out to be a realization of M-theory. That is, the future may reveal an analogy between the historical developments of quantum gravity and the Schrödinger and Heisenberg formulations of quantum mechanics. Were this to be the case, then one would expect fruitful new insights into all the issues raised here.

1.4 What quantum gravity and philosophy have to say to each other

Quantum gravity raises a multitude of issues interesting to philosophically minded thinkers. Physicists working in the field challenge some of our deepest assumptions about the world, and the philosophical tradition has a strong interest in many of these assumptions. In this final section we will describe a variety of topics in the subject that are of mutual concern to physicists and philosophers: some issues that bear on historically philosophical questions, and some new foundational issues. Our aim is to give philosophers a good outline of the problems that define the field, and to give physicists a sketch of how philosophers have investigated such issues – and of course to show how the arguments of our contributors fit into a broad dialogue concerning the foundations of quantum gravity. (Note that the papers are not exactly organized according to the following scheme, because many of them address several distinct issues.)

1.4.1 The demise of classical spacetime

- A number of the contributors make comments relevant to the 'fate of spacetime' in the quantum regime. Since we have already mentioned Witten's views, we will start there. He claims that, despite the original idea of strings propagating on a fixed background spacetime, spacetime arises entirely from the more fundamental two-dimensional conformal field on the string (an argument supported by duality symmetry). If correct (and if string theory is correct), then this view constitutes a considerable advance on our philosophical understanding of the nature of space. Views on the nature of space are as old as the idea of a general account of motion (see Huggett 1999): Plato and Descartes believed that matter and space were identical; Aristotle and other plenists often denied the existence of space by denying the vacuum of the atomists; and of course Newton and Leibniz were in famous opposition on the question of whether space was 'absolute' or 'relative'. The recent philosophical tradition (Friedman 1983 is especially influential here) has divided this question up in a number of ways: whether spacetime is dynamical or not; whether there is literally a manifold of points distinct from matter; and the meaning of 'relativity principles' as symmetries. Naturally, postgeneral relativity the answer to the first question is affirmative (though not without subtleties), but the other two topics are addressed by quantum gravity; the nature of the symmetries will come up later, for now we see the claim that in string theory spacetime is not distinct from matter (i.e. strings) but derives from it. The physical argument seems pretty straightforward, and it seems to us that philosophers should pick up this challenge: does Witten offer a sound answer to the problem of the nature of space, and how does it fit with historical proposals?

One oddity with this view is worth mentioning at once: the spacetime metric field that appears in the stringy Lagrangian must apparently be defined everywhere, not just on those points which the string occupies. Witten's spacetime seems to exist where matter and hence the fundamental two-dimensional field does not, so does it truely fade away?

- Canonical quantum gravity also has implications for the nature of spacetime, raised here by John Baez and Carlo Rovelli in Chapters 8 and 4 respectively. At first, it might seem that this approach should diverge from that of Witten by postulating a manifold as distinct from matter as that of general relativity, though subject to quantum effects: quantum spacetime should be to quantum matter as classical spacetime is to classical matter. However, as we shall mention below, it seems unlikely that canonical quantum gravity can be formulated in the space of 3-metrics as suggested earlier, and the significant advances in the theory of the last fifteen years have come from the 'loop' formulation due, *inter alia*, to Ashtekar, Rovelli, and Smolin, introduced here by Rovelli as a logical development of the insights of general relativity and quantum mechanics. The idea is that quantum gravity has as a basis (of quantum states) networks with spins values ($\frac{1}{2}, 1, \frac{3}{2} \ldots$) associated with the vertices. The picture may not seem intuitive, but if the nodes represent quantized regions, then suitable operators for geometrical quantities, such as volume, can be found, and so an understanding of the spacetime represented by such a state attained. (What is unknown at present is how the theory relates to general relativity in a classical limit.) If the loop basis is not unitarily equivalent to the 3-metric basis then (if it is correct) it too offers a picture in which spacetime is not fundamental, but a result of a more basic reality: in this case the spin network. Rovelli, in Chapter 4, claims that this too is a form of relationism.
- A closely related question is raised in Chapter 8 by Baez in his discussion of 'topological quantum field theory'. This is an approach to quantum gravity that gets away from a background spacetime by utilizing an analogy – brought out in 'category theory' – between the topological properties of a space and the quantum formalism. This enables one to construct models of quantum evolutions, involving topological change, satisfying a set of appropriate axioms, without worrying about the details of the dynamics of (quantum) geometry. The analogy is very suggestive, and may be a clue to uniting quantum mechanics and general relativity, but it comes at a price: the spacetimes in the theory have no local properties, such as a metric or causal structure, and so the theory cannot be the whole story. As Baez mentions, the ideas have been useful in canonical quantum gravity to study the dynamics in analogy to the Feynman approach of quantum field theory: one calculates sums over ways in which surfaces can interact by branching and joining – 'spin foams'.

An important philosophical question brought up here is how the causal structure of spacetime is to be built into this kind of theory. Causation is a perennial topic for philosophers: from Aristotle, to Hume and his sceptical descendants, who claim that causation is just 'constant conjunction' of some kind; to contemporary accounts in subjunctive terms – 'A caused B just in case if A *had not* happened then neither *would* B have happened' – or based on statistical considerations; to the understanding of causal structure in relativity. Apparently this is not a topic that is well understood in quantum gravity, but it is one that philosophers should address. Another question of philosophical interest here concerns the topology change that is the basis of the theory. Philosophers have rarely considered the exotic possibility of the topology of space changing with time, but it is a possibility with relevance to some traditional philosophical issues, for example Aristotle's, Descartes', and Kant's claimed 'unity of space' (see Callender and Weingard 2000).

- As we have just indicated, the fate of spacetime in many approaches to the subject is that classical spacetime structure (loosely, a semi-Riemannian metric on a continuous manifold) breaks down. In Chapter 2, Butterfield and Isham treat at greater length the consequences of the various programmes and speculations for the notion of spacetime. Common to these programmes and speculations is talk of the gravitational field in quantum gravity 'fluctuating'. But can this really make sense? In Chapter 3, Weinstein (a philosopher) takes this naive idea seriously, to see whether it can really hold water. He argues that it cannot without the theory failing to capture all observable gravitational phenomena. The main idea behind some of his critique is that fluctuations in the gravitational field imply fluctuations in the spatiotemporal, and hence causal, structure of the world. But it is hard to see how one can make sense of canonical commutation relations and hence quantize anything in the absence of a stable causal structure.

1.4.2 The nature of time

- Next, consider what is possibly the deepest of all philosophical puzzles, the nature of time (see, e.g. Le Poidevin and MacBeath 1993). Not only do certain programmes imply the breakdown of classical spacetime structure, but they also threaten to say something 'special' about time. Indeed, quantum gravity in all of its formulations seems forced to say something novel about this subject, for it must reconcile a conflict in the understanding of time between quantum theory and general relativity. Canonical quantum mechanics, since it is based on the Hamiltonian formulation, describes systems evolving with respect to a time parameter: either the preferred foliation of Galilean spacetime or – covariantly – the instantaneous hypersurfaces of an inertial frame. But general relativity is famously hostile to any such time parametrization (except in very special cases with nice symmetries). First of all, there is no such thing as time, *simpliciter*, in the theory, but rather a variety of time variables. There is the completely arbitrary co-ordinate time which, unlike time in Minkowski spacetime, has no metrical properties. There is also the proper time of an observer, but this cannot be extrapolated out to be the unique measure of time for all observers. And then there are various 'cosmic time' variables, such as Weyl's and Milne's definitions of cosmic time, though these are usually dependent upon special distributions of matter–energy or on special geometrical properties of the spacetime holding. In general relativity, there are plenty of cosmological solutions that do not even allow the possibility of spacetime being foliated by global spacelike hypersurfaces. Finally, the spacetime metric is dynamical in general relativity but non-dynamical in quantum mechanics. Thus, the two conceptions of time are very different. Quantum gravity, therefore, must say something about time, even if only that one of these two conceptions is fundamentally correct and the other only approximately right. Its verdict on this issue and others like it will be of interest to philosophers and physicists alike. In Chapter 2, Butterfield and Isham survey this problem (among others) and explain what some of the major programmes suggest is right.

- Some more specific problems connected with time arise in canonical quantum gravity. First, there is the notorious 'problem of time'. Since Belot and Earman discuss this in detail in Chapter 10, we will only briefly mention the problem. The problem is obvious enough: the Wheeler–DeWitt equation has no time dependence! Like a particle in a (non-degenerate) eigenstate of zero energy, the quantum state of the universe does not change, contrary to experience. (Note that this problem does not affect the classical theory: in that context a spacetime can evolve even if its Hamiltonian vanishes. It is only when we follow Dirac's prescription for the quantum interpretation of the constraints that trouble strikes.) Further, since the Wheeler–DeWitt equation must hold at all times, it holds before and after measurements. Thus there is no way to encode the information gained from a measurement back into the equation, which aggravates the first problem.
- Then, there is the related problem of observables. As we explained above, the natural interpretation of a gauge theory is that the gauge degrees of freedom do not correspond to physical transformations, and thus one concludes that all observable quantities must be gauge invariant, unchanged under the action of the gauge transformations (actually this is what is usually meant by a gauge symmetry). In formal terms this means classically that the Poisson bracket of any constraint and observable vanishes, $\{C, O\} = 0$; on quantizing according to the Dirac scheme, the operators corresponding to the constraint and observable must commute, $[\hat{C}, \hat{O}] = 0$. Now in general relativity we find that the Hamiltonian, H, which is supposed to generate evolutions is itself a constraint, and so commutes with any observable; but $\{H, O\} = 0$ and $[\hat{H}, \hat{O}] = 0$ mean that any observable quantity (its value or expectation value) is a constant of the motion. Equivalently, since the two parts of the Hamiltonian generate diffeomorphisms, all physical quantities must be diffeomorphism invariant. As a rule, constants of the motion tend to be pretty dull physical quantities. In fact, if Σ is compact, the system has no known observables; if Σ is open, then trivial quantities may be defined, but they are generally acknowledged to be useless to quantum gravity. In particular, since the spatial 3-metric presumably changes with time, it cannot be an observable (which leaves one wondering what the point is of promoting it to an operator). Of course this flies in the face of the conventional understanding of general relativity, which holds that the metric is observable.

  Note that although the problem described afflicts both classical and quantum versions of the theory, it raises an additional conflict between quantum general relativity and the usual interpretation of quantum mechanics. Any quantum mechanical observable with time-dependent values will satisfy $[\hat{H}, \hat{Q}] \neq 0$, which means that there are no simultaneous eigenstates of $\hat{H}$ and $\hat{Q}$. But since H is a constraint, the fundamental postulate of Dirac's quantization is that every state of the system is an eigenstate of $\hat{H}$: $\hat{H}\Psi = 0$. But then no possible states are eigenstates of such a $\hat{Q}$, and so the usual account of quantum measurement breaks down: there simply cannot be a collapse into a state of definite Q if no such states exist!

Proposed solutions to these problems are compared and evaluated in detail by Belot and Earman in Chapter 10, and we think they are only half-joking when they divide them into 'Parmenidean' and 'Hericlitean' kinds. Parmenides (later followed by the more familiar Zeno) argued that all change was illusion, and the view associated with Rovelli (intimated in Chapter 4, but outlined more fully in Chapter 10 by Belot and Earman) has this character. According to Rovelli, all physical quantities are irreducibly relational, and hence timeless: for example, 'the clock read one as the mouse ran down' is a physical property, but it is not analysable into '(at t the clock read one) and (at t the mouse ran down)', since readings and positions at t are not physical. The idea of course is to bite the bullet, and accept that only diffeomorphism invariant quantities are physical, so there are only timeless truths.

Heraclites, on the other hand, argued for permanent flux, and the views proposed by Kuchař (e.g. 1992), among others, have this character. In the more extreme formulations they go against the spirit of general relativity and assume a preferred foliation, which gives physical significance to observable quantities that are time-specific. Kuchař's proposal is more subtle, seeking to find a meaningful time-parameter without violating the spirit of general relativity, but strong enough to deny that the scalar constraint has the same force as the vector constraint: quantities should be invariant under spatial diffeomorphisms, but not under timelike diffeomorphisms.

We would speculate that the problem of change is perhaps the oldest philosophical subject, and its solutions are the source of many metaphysical problems. The oldest version asks how change is possible at all: if A changes, then it is no longer the same, and hence no longer A, so A has not changed, but ceased to exist! In recent years the problem has been most focussed on the nature of identity – especially the nature of persistence through time – and the meaning of the spacetime view of the world: are all truths 'tenseless', or is some sense to be made of a 'specious present' (see Le Poidevin and MacBeath 1993 and references therein; Williams 1951 is an enjoyable classic article on this topic). Clearly, the views of Rovelli and Kuchař have crucial bearing on these arguments, and must be taken seriously by philosophers of time, though, as Belot and Earman point out in Chapter 10 , which (if either) of their insights is correct is something to be determined in part by the success of their approaches in solving physical problems.

1.4.3 The interpretation of general relativity

• Since diffeomorphism invariance is the symmetry underlying general relativity's general covariance, the interpretation of general covariance – a hotly disputed topic since the theory's inception – may be relevant to solving some of the above problems. Indeed, as many of the chapters in this volume make abundantly clear (especially Chapters 9 and 10), there most certainly is a connection between issues in quantum gravity and the interpretation of general covariance, and in particular, Einstein's famous hole argument, which itself bears on the relational–substantival debate concerning space. In recent years this argument has attracted a great deal of attention among philosophers. Earman and Norton (1987) argued that the example shows that the 'manifold substantivalist' – one committed to the literal existence of a spacetime manifold distinct from matter – had to accept radical indeterminism: there are infinitely many models of general relativity that agree outside the hole, but disagree inside. Put another way (as Belot and Earman suggest in Chapter 10), the substantivalist must accept that the points of two manifolds can be compared, independently of the various fields on them; that is the meaning of their distinctness from matter. Substantivalists have replied in a variety of ways, some by arguing that while the comparison can be made without regard to matter fields, it cannot be made without regard for the metric field, others use the developments in our understanding of the logic of possibility to argue that diffeomorphism invariance is compatible with substantivalism. Since one's preferred solution to the problem of time bears on the meaning of diffeomorphism invariance, it is only natural to expect that it will have bearing on the hole argument and hence the nature of spacetime. (Looked at from another angle, as Belot and Earman again point out, the issues here are a very special case of the philosophical issues that arise in understanding gauge degrees of freedom, which are at once unphysical, but also seemingly essential for modern field theories. See also Redhead 1975.)

- Another issue, raised by Penrose in Chapter 13, suggests that there are problems with approaches in which there are quantum superpositions of spacetime, as, for instance, there would be in the intuitive approach to canonical quantum gravity we sketched in Section 1.3 (though it is less clear what to say about the loop approach). Penrose points out that general relativity's principle of general covariance seems at odds with the quantum mechanical principle of superposition. For suppose $|\alpha\rangle = |\phi(x)\rangle|\Psi\rangle$ represents a state in which a particle is localized in a spacetime with a sharp metric, and $|\beta\rangle = |\phi(x + a)\rangle|\Psi'\rangle$ represents the particle shifted and the appropriate new metric eigenstate. Now, we can imagine the state $|\alpha\rangle + |\beta\rangle$ representing an entangled system involving particle and gravitational field superpositions. Certainly such a state is distinct from either $|\alpha\rangle$ or $|\beta\rangle$, but how can these two states be distinct? Since the particles only differ by a displacement, we can suppose the metrics to as well, and so the spacetimes involved are diffeomorphic: given (one reading of) general covariance the two spacetimes and hence quantum states are one and the same.

  In Chapter 9, Julian Barbour tackles this issue (and the preceding one), arguing that there is a canonical way to identify points between slightly differing spacetimes; the key insight, he thinks, comes from a method of deducing the 'best match' between one relative configuration of particles and another. In the case of 3-geometries he claims this method is simply Hilbert's variational principle. This idea traces its origins to work in mechanics by Lagrange, Lange, and Mach, among others, and it has relevance to Machianism about spacetime and the hole argument.
- It is often thought that only quantum mechanics has a problem of interpretation. However, the above issues and others make it plausible that even general relativity has a problem of interpretation. The interpretation of general relativity is of

obvious importance to quantum gravity: it is important to understand properly the assumptions that go into relativity so that one can better appreciate what one can and cannot give up when creating a new theory. In this spirit, in Chapter 11 Harvey Brown and Oliver Pooley examine John S. Bell's (1976) paper 'How to Teach Special Relativity', and apply the lesson they learn from it regarding special relativity to general relativity. Bell's paper explains the Lorentz transformations dynamically à la Lorentz: by showing how they can be derived from the structure of matter (in particular from electrodynamics) in relative motion. Einstein related this explanation of relativity to his own account, as kinetic theory explanations (so-called 'constructive theory' explanations) relate to thermodynamical explanations (so-called 'principle theory' explanations). Like Bell, Einstein claimed that Lorentz's constructive explanation was necessary for a proper understanding of special relativity. Brown and Pooley seek to extend this point to general relativity. They claim it clarifies the role of kinematics and dynamics in special and general relativity, as well as the role of rods and clocks in the two theories.

1.4.4 The interpretation of quantum mechanics

- Another connection between philosophy and quantum gravity involves the notorious measurement problem in quantum mechanics. Physicists and philosophers of science have both devoted much time and energy to discussing this topic. Is the measurement problem and the interpretation of quantum mechanics relevant to quantum gravity? In this volume we are fortunate enough to have sharply divided answers to this controversial question. In Chapter 4, Rovelli emphatically claims that there is no connection between the two, whereas Sheldon Goldstein and Stefan Teufel in Chapter 12 claim that the connection makes all the difference in the world, and that the failure to acknowledge it is responsible for many of the conceptual problems in the field, such as the problem of time.
- The Bohmian approach, here advocated by Goldstein and Teufel, also falls under this heading. In the case of particle quantum mechanics, Bohm's theory describes point bodies – 'beables' – whose definite motions are determined by their collective locations (not momentum) and the wave function, which is itself determined by the ordinary Schrödinger equation (not the particle locations). Goldstein and Teufel explain in Chapter 12 how this picture can be carried over to canonical quantum gravity, where the evolution of a definite 3-geometry – the beable – is determined by a wave function, which is itself determined by the Wheeler–DeWitt equation. This time, since the wave function is not representing the physical state but, as it were, driving the 3-geometry, the problem of time does not arise: there is no inconsistency in a stationary wave function leading to an evolving spacetime in this theory. Further, in Bohmian mechanics the usual quantum formalism for observables is not taken as a matter of fundamental postulate, but rather should emerge from analysis of experiments within the framework. In this case the problem of observables cannot get off the ground: whatever quantities can be shown to be measurable in experiment are observable. Finally, since physical quantities come directly from the beables, not the wave function, and since in a model of Bohmian quantum gravity there is only one stationary wave function, it also seems

that the need for an inner product is moot: the theory simply sidesteps one of the hardest technical problems of canonical quantum gravity (explained below). On the downside, despite many interesting applications of Bohmian quantum gravity to conceptual and physical problems – for example to the problem of time by Holland (1993), Callender and Weingard (1994, 1996), and many others; to black holes by Kenmoku et al. (1997); to the initial singularity by Callender and Weingard (1995); and to various cosmological models by Blaut and Kowalski-Glikman (1996) – Bohmian quantum gravity has not been developed as seriously as some other approaches, so it is difficult to say whether or not it has hard problems of its own.

- In Chapters 13 and 14, Roger Penrose and Joy Christian also see a connection between gravity and the measurement problem. Penrose develops the attractive idea that the gravitational field can be drafted in to help answer the measurement problem. The idea is that the gravitational field is the 'trigger' that stimulates collapses of quantum superpositions before they become macroscopic. In principle, this interpretation should give slightly (though currently unobserved) different results than standard quantum mechanics. We therefore have the exciting possibility of testing this collapse theory against the standard theory. Penrose, here and elsewhere, proposes some such experiments. But will they really succeed in testing his theory? Christian claims that one type of experiment will not work and proposes some others in its place, as well as greatly elaborating the conceptual position of Penrose's model within quantum gravity.
- Issues analogous to those concerning collapses arise from Hawking's (1974) celebrated results concerning the 'evaporation' of black holes (the most important achievement of the semiclassical theory). In a non-unitary wave function collapse, information about the prior state must be erased: one cannot invert the dynamics to reconstruct the original wave function. Similarly, information about a system's state falls into a black hole with the system, but – on the standard treatment – is neither contained in the thermal radiation given off as the black hole evaporates away nor in the black hole itself. Like measurement, it is hard to reconcile this 'information loss' with the unitary evolution required by quantum mechanics.

  In Chapter 7, Unruh explains the ideas of black hole thermodynamics, and especially the significance of black hole entropy. He also uses an ingenious parallel between light transmission in spacetime and sound transmission in moving water (if a black hole is a region from which light cannot escape, then a 'dumb hole' is a region from which sound cannot escape, for instance because the water inside is moving in the opposite direction supersonically) to show that Hawking's result is insensitive to short length scale physics. However, it is such short length physics, and in particular string theory, that may offer a solution to the information loss problem according to recent proposals.

  First of all, one can (in some very special cases) calculate a 'traditional' entropy for a black hole by counting the number of corresponding (in a loose sense) string states and taking their logarithm; the result is (in those special cases) in agreement with the Hawking-style calculation. This raises the possibility that the information about systems which fall into the black hole is in fact contained in the strings.

That is, knowing the details of the thermal radiation and the exact string state might suffice to reconstruct the in-falling state, as knowledge of the radiation and internal state of a hot poker would allow resconstruction of the heating process. Unruh criticizes this proposal, arguing that if the strings are inside the black hole then they cannot influence the thermal radiation in the appropriate way, but if they are outside the black hole then they cannot be suitably affected by the in-falling system. The only possiblity seems to be some kind of non-local interaction between strings inside and outside the hole, but this scenario is deeply unappealing.

1.4.5 The status of the wave function

- Another topic of interest to both physicists and philosophers concerns the status of quantum cosmology. Quantum cosmology, in contrast to quantum gravity, aims to provide a rationale for a particular choice of boundary conditions for our universe. The most familiar schemes of this kind are the famous Hartle and Hawking (1983) No-Boundary Proposal and the Vilenkin (1982) initial wave function of the universe. Some physicists conceive of quantum cosmology as a prescriptive enterprise: they believe that there are laws of quantum cosmology. Hawking, Hartle, and Vilenkin are all engaged in what we might call 'cosmogenic' theories. They are trying to find laws that uniquely determine the initial conditions of the universe. But is this search scientifically respectable? What possible justification could there be for the choice of a particular boundary condition – aside from the fact that it works, i.e. that it leads to what we observe? Any inductive inference from a single case is unwarranted, so how can we scientifically justify talk of laws and causes for the universe as a whole? While none of the contributors addresses this question directly, Goldstein and Teufel in Chapter 12 do speculate about the meaning of the universal wave function. And these considerations do raise the question of whether the usual probabilistic interpretation of the wave function can be carried over to the case of the universal wave function.
- Related problems arise in particular approaches to quantum gravity. To give an example, consider the problem of the interpretation of the Wheeler–DeWitt equation (eqn. 1.8). The naive suggestion is that it receive the same interpretation as does quantum mechanics. (By interpretation we here mean only rules for extracting predictions, not a solution to the measurement problem; until we know how to extract predictions from the theory it doesn't even have the *luxury* of having a measurement problem!) The naive interpretation would be to think of $\Psi(h, p)$ as a probability, which when squared yields the probability that an observer will measure the values h and p. Even putting aside the question of where this observer of the whole universe is, we know this scheme cannot work (at least straightforwardly). This can be seen by comparing the Wheeler–DeWitt equation to the more familiar Klein–Gordon equation. If we impose the Klein–Gordon Hamiltonian as a restriction on the space of physical states, then the analogy between the resulting equation and the Wheeler–DeWitt equation is very strong. In fact, when the Wheeler–DeWitt equation is reduced to two degrees of freedom, it *is* this resulting equation. Now recall that the Klein–Gordon equation suffers from a very serious problem: its inner product $\langle\Psi_1|\Psi_2\rangle$ is not positive definite and therefore cannot be

used to define a probability. Thus we have a problem with defining a sensible inner product for the position representation (see, e.g. Teller 1995). The same problem threatens here, and is one motivation for loop quantization, for one can define an inner product on the space of 3-metrics in that approach.

The first sections of this introduction sought to clarify the motivations for the search for a theory of quantum gravity and give a useful outline of the two major programmes. In this final section we provided a sketch of many (but surely not all) of the major philosophical dimensions of quantum gravity. We hope that the reader, equipped with these ideas will now have sufficient context to tackle the chapters of this book, seeing how they relate to the broad physical and philosophical issues that surround the topic; if so, he or she will surely find them as exciting and illuminating as we have. We look forward to the debates that they will spark!

Notes

Many thanks to John Baez, Jossi Berkovitz, Jeremy Butterfield, Carl Hoefer, Tom Imbo, Jeffrey Ketland and Carlo Rovelli for indispensable comments on this introduction. Portions of the Introduction are based on 'Why Quantize Gravity (or Any Other Field for that Matter)?', presented at the Philosophy of Science Association Conference (Callender and Huggett 2001).

1. In fact, it is these assumptions of classicality that do the work in the argument; the fact that the field of interest is not just classical, but also gravitational, does not play a role.
2. Obviously this thought experiment relies on some extreme idealizations, since we in effect postulate that the electron is screened off from all other fields so that no correlations are lost. But this does not detract from the point of the example: superluminal signalling that picks out a preferred foliation of spacetime must be impossible in principle, not just practice, if we take relativity seriously and literally.
3. A couple of other comments: (a) If Lefty found her box empty then Righty would still measure an effect, providing that the scattering is sensitive to the amplitude of the wave function (if not, one could still arrange signalling either by using enough boxes to ensure that Lefty will find an electron in one, or by using one of the other schemes mentioned above); (b) As Aharanov and Vaidman (1993) point out, this kind of arrangement will not permit signalling with their 'protective observations', though they are similar in allowing measurements of the wave function without collapse.
4. Of course, one might be concerned that such *ad hoc* fine-tuning of parameters is an indication that we are reaching a 'degenerating' phase of the spontaneous localization program, adding epicycles to save the theory. But this judgement may be premature: Pearle and Squires (1996) suggest how some such parameters may be derived.
5. Alternately, one might try to maintain momentum conservation *on average* (as Brown and Redhead 1981, footnote 21, suggest) in collapses. That is, one could seek to complete the collapse dynamics of QM in such a way that the expectation value for momentum was always conserved.
6. If gravitational waves do cause quantum jumps, then a wave empinging on the Earth could clearly have disastrous consequences if it were sufficiently powerful to collapse every piece of matter!
7. One issue that has captured a great deal of attention in the semiclassical theory is the so-called 'loss of information' problem (e.g. Belot, Earman, and Ruetsche 1999): as the black hole evaporates away, there is a transition from a pure to mixed state for the matter fields, reminiscent of the collapse in measurement. While this has worried many, it is not really so surprising given what we have been saying: no unitary evolution can produce such a transition, but in the model one is effectively invoking eqn. 1.4 as a second – non-unitary – equation of motion. It is really just another reflection of the point that in a half-and-half approach, one must be careful about how to include collapses.

8. So too does comparison with our toy example of a constraint. For instance, if one takes electromagnetism written in terms of the gauge potential A, and attempts to remove the unphysical degrees of freedom by reparameterizing with gauge invariant quantities, such as the magnetic field, $B = \nabla \times A$, one introduces non-locality into the theory (as shown by the notorious Aharonov–Bohm effect).
9. We should point out that there are reasons to wonder whether this intuitive picture of 3-metrics evolving on superspace is anything other than a metaphor. The 3-metric does not weave its path between Cauchy hypersurfaces like point particles in classical dynamics do. Wheeler's so-called 'Thick Sandwich' conjecture is false: that given any two 3-metrics in superspace there exists a spacetime between them such that they arise as induced metrics on two disjoint Cauchy surfaces. Different foliations between the two surfaces changes the curve between them. And even the Thin Sandwich conjecture – that given a point and tangent vector in superspace, there is a unique spacetime realizing the initial condition – has only limited applicability. See Bartnik and Fodor (1993) for more.

   On the positive side, note that Christian (1997) shows how to exactly quantize a simpler spacetime theory – Newton–Cartan theory – along these lines.