Worlds in the Everett interpretation

David Wallace*

This is a discussion of how we can understand the world-view given to us by the Everett interpretation of quantum mechanics, and in particular the rôle played by the concept of 'world'. The view presented is that we are entitled to use 'many-worlds' terminology even if the theory does not specify the worlds in the formalism; this is defended by means of an extensive analogy with the concept of an 'instant' or moment of time in relativity, with the lack of a preferred foliation of spacetime being compared with the lack of a preferred basis in quantum theory. Implications for identity of worlds over time, and for relativistic quantum mechanics, are discussed.

Keywords: Interpretation of Quantum Mechanics — Everett interpretation; Preferred Basis; Decoherence; Spacetime Foliation

1 Introduction

If we take unitary quantum mechanics seriously *á la* Everett (1957), the view of reality which emerges is certainly not that of a single classical universe. But is it a 'many-worlds' view (as DeWitt (1970) and Deutsch (1985) have argued)? We can object to this view *taken literally* on many grounds: ontological extravagance, difficulties with relativistic covariance, violation of the principle of the identity of indiscernibles, the need to find an exactly preferred basis, etc.[1] But if we reject such a view then how are we to understand the Everett interpretation — what metaphysical picture does it give us?

Looked at another way, defenders of an Everettian viewpoint face a dilemma: just how seriously are we to take these worlds? On the one hand, if we were to take them literally and build them into our formalism then we would face the preferred-basis problem at its worst. Decoherence would be no use to us, for we would need an *exact* world-defining principle and not some pragmatic criterion. Furthermore, there seems to be no relativistically covariant way to define a world, so this approach would push us towards the same problems with

[1] See Barrett (1999) for details of these criticisms.

(*March 15, 2001*)

* Centre for Quantum Computation, The Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, U.K.
(*e-mail:* david.wallace@merton.ox.ac.uk).

relativity faced by approaches like the pilot-wave theory (Bohm 1952; Holland 1993) or state-reduction theories (Ghirardi, Rimini, and Weber 1986; Pearle 1989). On the other hand, if we banish these worlds from our formalism we must answer the criticism that our theory is just uninterpreted mathematics, and explain how sense can be made of the universal state.[2]

In this paper I shall attempt to defend three claims. The first is that Everettians can avoid this dilemma by a compromise: they may legitimately and meaningfully use the terminology of many worlds without being required to represent these worlds in their formalism. The second claim is that this view of Everett, rather than being either contradictory or a pathology confined to quantum mechanics, is typical of what we should expect when interpreting a theory which changes our concept of what the world is. The third claim is that developing such a view of quantum mechanics brings out a number of deep analogies with the other twentieth century physical theory which radically changed our concept of the world: relativity.

The second of these claims is defended in section 2, in which I shall argue that any theory which revises our previous ways of understanding reality nevertheless needs to make contact with these previous theories if we are to be able to comprehend the new theory. These previous ways are wrong in some respects (else we would not need the new theory!) so their use in an interpretation of the new theory may be incomplete or obscure; nonetheless such use is essential if the new theory is to be seen as a theory of *physics*. We may hope eventually to come to understand a given theory without describing any of its concepts in this way, but we cannot begin without any reference to what has gone before.

In section 3 the classical concept of relativistic spacetime is used as a concrete example of this process, but the main reason for discussing spacetime extensively is to set the stage for a discussion of the profound analogies between (relativistic) spacetime and the Everettian universal state. These analogies have already been discussed — notably, and extensively, by Simon Saunders (1993, 1995, 1996b, 1997, 1998) — in the context of the probability problem in Everett interpretations.[3] My approach is complementary to this (and I mention probability only in passing): if we accept the Everett interpretation, we may use the spacetime analogy to cast light on how we can describe and understand the world-view it gives us. (A sort of converse also holds: if we insist that the Everett interpretation is meaningful only with a preferred basis then we are forced to accept that relativistic spacetime is meaningful only with a preferred foliation.)

In the remainder of the paper I shall attempt to sketch out this approach, and in doing so to defend my first and third claims. In section 4 I address

[2] In either case, of course, we must explain how to understand probability; but that problem lies beyond the scope of this paper.

[3] The problem here is (at least in part) how to do without a criterion by which we can individuate worlds and identify how they change over time, or how to provide one if it turns out to be indispensable. Saunders, and Tappenden (2000a, 2000b), have argued that such a criterion is not needed, and this view will be adopted here (see however Barrett (1999) for a dissenting view, and Butterfield (1996) for further discussion).

the problem of how we make sense of the Everett interpretation without an exactly preferred basis, while in section 5 I discuss the question of the persistence of worlds through time and in section 6 I discuss the role of the observer in the Everett interpretation, and contrast the approach of this paper with many-minds approaches. In sections 7–9 I consider how we are to understand (special-)relativistic quantum theory, and in sections 10 and 11 I summarise the approach.

2 Interpreting our Theories

What is spacetime? A mathematically motivated answer (see, e. g., Wald (1984)) might go something like

> Spacetime is (or possibly, 'is isomorphic to') a set $\{\mathcal{M}, \mathbf{g}, \phi_1, \ldots \phi_n\}$, where $\mathcal{M}$ is a connected smooth paracompact 4-manifold, $\mathbf{g}$ is a smooth symmetric 2-tensor field on $T\mathcal{M}$, and the ϕ_i are further tensor fields on $\mathcal{M}$.

But *given* such a mathematical description, how are we to understand what spacetime is? Assuredly we have to get such an understanding, for without any way to interpret the mathematics as a description of the world our theory can have neither predictive nor explanatory power. After all, suppose that we had instead said that

> Spacetime is a set containing exactly six elements, together with a commutative group operation on that set of elements.

To object that this is nothing like spacetime, we need some inkling of what sort of object this 'spacetime' actually is.

This sort of requirement is general to mathematical physics. To interpret a mathematically-formulated theory as physics we need some way of linking the mathematics to the physical world. At the least, this requires that we locate *ourselves* somewhere in the theory, or else that we have some idea of how the objects and concepts of our everyday world are mapped into the mathematics. We are able to accept that, in the light of new theories, these objects and concepts may turn out to be not exactly what they seem — and indeed some may be completely illusory — but our theory needs at least to make contact with them in order for us to do physics.

It is, however, important to remember that 'our everyday world' is not something to which we have direct sensory access. The existence of a three-dimensional spatial universe in which exist various macroscopic objects — including ourselves — and which we learn about through vision, touch etc. is, rather, an extremely effective theory (the 'everyday theory'), which explains and unifies our observations. This point is familiar from Quine's writings (see, e. g., Quine 1966).

At first sight, it then seems that to interpret theories of physics by finding the everyday world within them is to court infinite regress, for we would seem required in turn to interpret the 'everyday theory' in terms of something yet more basic. (After all, once we have accepted it *as* a theory we may formalise it in mathematical language every bit as forbidding as that used to describe spacetime.[4] As such, if we are required to explicate spacetime then why not our theory of space?) It might appear that we would instead do better to link our new theory directly to our observations — presumably if the new theory is an improvement upon the 'everyday theory' it too will explain our observations, only better.

The problem with this idea is that our observations are themselves inextricably entwined with the theory of the everyday world. To take vision as an example,[5] it is a gross over-simplification to suppose that our visual field is simply a region of coloured patches rather like a television screen. Much low-level processing of visual data takes place before it is presented to higher-order subsystems of the brain and begins to affect our actions (we might say 'before we are conscious of it'). For a start we have two eyes, and the parallax calculations that tell us the distance to nearby objects appear to be made at a fairly early stage of processing. We do not consciously work out distances to objects, we are simply aware of their distances in addition to their angular positions. But there is far more processing than this: identification of objects, making assumptions about regions of the visual field that cannot be observed (such as the blind spot) etc., so the information presented is probably better described along the lines of "there's a book-case at 3 o'clock, 4 metres away to judge by the parallax, five shelves, lots of multi-coloured books, all about thirty centimetres high" than by a pixel-by-pixel description of the bookcase and the rest of the visual field.

So there are no theory-neutral observations: rather, there is an existing theory in terms of which our observations are automatically interpreted, and which we must take as our starting point when interpreting the theories of physics.[6] It should come as no surprise that we are capable of comprehending

[4] "Everyday space is a smooth three-manifold together with a smooth symmetric two-tensor $\mathbf{g}$, whose unique associated metric-compatible symmetric connection has zero holonomy, ...".

[5] This discussion is based on material taken from Dennett (1991b).

[6] This is not a novel view; for an example of a similar viewpoint, consider Einstein (reported in Heisenberg (1971, pages 63–64)):

> It is quite wrong to try founding a theory on observable magnitudes alone. In reality the very opposite happens. It is the theory which decides what we can observe. You must appreciate that observation is a very complicated process. The phenomenon under observation produces certain events in our measuring apparatus. As a result, further processes take place in the apparatus, which eventually and by complicated paths produce sense impressions and help us to fix the effects in our consciousness. Along this whole path—from the phenomenon to its fixation in our consciousness—we must be able to tell how nature functions, must know the natural laws at least in practical terms, before we can claim to have observed anything at all. Only theory, that is, knowledge of natural laws, enables us to deduce the underlying phenomena from our sense impressions. When we claim that we can observe something new, we ought really to be saying that, although we are about to formulate new natural laws that do not agree with the old ones, we nevertheless assume that the existing laws—covering the whole path from the phenomenon to our consciousness—function in such a way that we can rely upon them and hence speak of "observation."

this theory directly, for we are evolved to live in a universe described reasonably well by it, and so it is hardwired into our brains. New theories of physics do not have this advantage: perhaps through long exposure to them we might come to have a similarly direct grasp of them (rather than understand them only via the existing theory) but this may not be possible at all and certainly will not be possible when we are first presented with the new theories.

To summarise, then, our program for interpreting a new theory is inevitably dependent on an existing theory whose interpretation we are assumed to understand already (such an existing theory will either be our everyday theory of bodies, or will in turn be interpreted by means of that theory). The requirement is that certain parts of the new theory must be 'approximately isomorphic' to the existing theory.[7] The isomorphism may be imperfect, revealing parts of our old theory to be in error, and our understanding of the old theory may be greatly transformed, but it seems that a recognisable shadow of the old theory must exist in the new one for it to be more than just mathematics.

(For reasons of space I shall not develop the links between this material and the closely related questions of how we are to understand theory change and in what sense one theory can be contained within another. The approach adopted here, however, has much in common with the structural realist program in the philosophy of science (see Worrall 1989; Psillos 1995; Ladyman 1998).)

3 Relativistic Spacetime and the 'Many-instants' Interpretation

To illustrate this view of theories, let us return to the question raised at the start of the last section: how are we to understand (Minkowski) spacetime?[8] One natural way to do so makes essential use of the concepts of space *and* time, regarded initially as separate. One can say of any given event that it is in a given place, at a given time — the former property being represented by a point in three-dimensional space, the latter by a single number. As a matter of mathematics, then, we can construct a four-dimensional 'space' (using the word in its mathematical sense) which consists of all points in space, at all times. This space is a stack of the various spaces each of which is space as we ordinarily understand it, at a given time. This might be called a 'many-instants' interpretation of spacetime: spacetime is the collection of all instants of time, together with a metric structure connecting them and representing the time elapsed between instants.

On understanding better this metric structure — specifically, on recognizing its Lorentzian symmetries — we see that this description of spacetime obscures its nature in many ways. For a start, it is clear that any given way of slicing it up is arbitrary, and that many other choices of simultaneity could be made. (In more mathematical language we are saying that given a relativistic spacetime,

[7]The reader may, if desired, interpret this as saying that models of the existing theory must be approximately isomorphic to parts of models of the new theory.

[8]Throughout this paper, 'spacetime' means Minkowski spacetime, though most of what I have to say applies to any relativistic spacetime. When quantum field theory is discussed, however, the spacetime metric will be assumed to be fixed and classical: quantum gravity will not be discussed.

we can make sense of what that spacetime is by considering a global foliation — but that no such foliation is the preferred choice.) This becomes worse when we start to interpret simultaneity in the spirit of general relativity: that is, entirely as an artificial construction. We then realise that there are a multitude of spacelike slices, identical in our vicinity, which differ from one another at greater distances.

The situation seems to worsen when we remember the theoretical nature of our concept of space. After all, the surface we perceive through our eyes is not spacelike at all: it is the past light-cone, which does not bear much conceptual resemblance to our ordinary concept of space (in particular, it has wholly the wrong causal structure). And furthermore, our everyday understanding of 'space' comes with the understanding that we are represented in it, that our conscious thoughts supervene on some region of space. But barring radical dualism, our thoughts presumably must supervene on some processes in our brains — and processes take finite time to occur, and occupy finite space.[9] Hence our thoughts supervene on regions of spacetime — so if we regard spacetime as a collection of instants, we have to face up to the fact that a single conscious thought supervenes on many instants ('many moments of time') rather than just one.

So, as a description of spacetime the many-instants interpretation leaves much to be desired: it contains arbitrariness, obscures structural features, and we ourselves cannot be said to exist in any single given instant. Nonetheless it is a description with merits. Specifically, it is a *complete* description: once we have specified all the contents of all the instants of time and the temporal relations between instants, we have the entire spacetime. Further, we have an existing intuitive grasp of what an instant of time is like, and from this we can gain some understanding of what sort of entity spacetime is. This understanding is enough to make contact between our experiences and the formalism of relativity, granting the theory predictive and explanatory power.

Furthermore, for everyday purposes (i. e. when we confine ourselves to re-

[9]I regard this point as uncontroversial. I do not claim that this finite extent of (the object of supervenience of) our thoughts is of the order of the 'specious present' — it may well be that the 'finite time' of these processes to be measured in nanoseconds — but if we want to hold to any sort of functionalist view of the brain then we're going to be stuck with *some* finite duration.

I suppose it's possible, given our limited understanding of the nature of consciousness, to imagine that our thoughts actually supervene on instantaneous states of the brain, rather than processes. But this seems highly implausible: the brain still has nonzero spatial extent so the objects of supervenience would have to be some specific 3-slices of the world-tube of the brain, and it is hard to imagine a plausible method to describe *exactly* which slicing to choose — eventually any method based on, say, Lorentz frames will need to make arbitrary choices of exactly how we define the brain, and in any case it seems hard to motivate such an exact rule. It might be more plausible for the objects of supervenience to be only approximately defined, in which case they could perfectly well be states of the brain — after all neurologists talk perfectly sensibly about 'the current state of a brain' without worrying about the difficulties of defining *precisely* the choice of simultaneity hypersurface which they wish to use — but in a certain sense we'd still supervene on many instants at once. If we do wish to defend a claim of supervenience on an instant, a substantial metaphysical package would seem to be required — such as that developed by Barbour (1999).

gions of modest size and modest velocities) there is a pragmatically preferred foliation, a pragmatically preferred way to regard spacetime as a collection of instants: that foliation defined by the Lorentz frame appropriate to our current state of motion. Granted that it is irrelevant whether at large distances we choose to let the foliation deviate from the Lorentzian choice; granted that if we look closely enough we will need to make arbitrary choices as to whether to take a Lorentz frame appropriate to, say, the velocity of Oxford or of a plane flying over it. For everyday purposes the preferred foliation is effectively well-defined. It is here, in the elements of this foliation, that we find those entities approximately isomorphic to the three-dimensional spaces of our everyday world-view.

Note here how important it is that we identify our everyday view of the world as a theory: if we did see ourselves as having direct access to it we would have to map it onto the past light-cone and not some spatial hypersurface. As it is, though, we can claim to have identified ordinary space (approximately) in our new theory, which tells us how in the new theory to describe the universe we observe and how to locate ourselves in the theory.

To say that spacetime

> is an entity with such-and-such mathematical description *which can be thought of as a collection of all instantaneous spaces together with the metric relations between them, no matter that it may be decomposed into such a collection in many different ways*

is an inelegant way of describing it. But it is an accurate one, and since Nature and our early education have not seen fit to provide us with a direct intuitive grasp of what spacetime is we must begin with such inelegant ways to understand it. Ideally we might, through working with the spacetime concept, develop a direct understanding of it, in which case this ladder can be kicked away. But I doubt that we are yet sufficiently at ease with the concept to allow this.

4 Quantum Theory

The connection with Everett should be fairly clear. Everett's essential postulate is that the quantum state, and its unitary evolution, are universal; but this leaves us with another forbiddingly unintuitive mathematical description (Haag 1996):

> The universal state is a set $\{\mathcal{H}, \mathcal{M}, \widehat{\phi}_1, \ldots \widehat{\phi}_n, \rho\}$ where $\mathcal{H}$ is a separable Hilbert space, $\mathcal{M}$ is a spacetime, the $\widehat{\phi}_i$ are (distributional) maps from $\mathcal{M}$ to the algebra of bounded operators on $\mathcal{H}$, and ρ is a self-adjoint bounded operator on $\mathcal{H}$ of trace 1.

Without some idea of how to interpret this entity we clearly cannot treat it as physics.

In trying to understand what the Everett interpretation implies we typically begin in the middle, presuming various concepts already understood. We might for instance begin with Schrödinger's cat, and suppose some subsystem of the universe to be described by a Hilbert space $\mathcal{H}$ which is the tensor product of spaces describing a radioactive atom, a Geiger counter, a cat, and some auxiliary apparatus (the box, the cat-killing device, etc.). We also suppose that there are some states in these component spaces which represent definitely decayed or undecayed atoms, triggered or untriggered counters, dead or alive cats, and so forth.

Note how extensively we use the concepts of our previous theory. In particular we suppose that we know what is meant by an alive or dead cat, or a triggered or untriggered counter — it is superpositions of systems in such states that we want to understand.

Following the steps of Schrödinger's thought experiment, we know that the unitary dynamics will lead to some evolution like[10]

$$\begin{aligned}
&\left|\begin{array}{c}\text{atom}\\ \text{undecayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter}\\ \text{untriggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison}\\ \text{in vial}\end{array}\right\rangle\left|\begin{array}{c}\text{cat}\\ \text{alive}\end{array}\right\rangle \longrightarrow \qquad (1)\\
&\frac{1}{\sqrt{2}}\left(\left|\begin{array}{c}\text{atom}\\ \text{decayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter}\\ \text{triggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison}\\ \text{released}\end{array}\right\rangle\left|\begin{array}{c}\text{cat}\\ \text{dead}\end{array}\right\rangle +\right.\\
&\qquad\left.\left|\begin{array}{c}\text{atom}\\ \text{undecayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter still}\\ \text{not triggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison still}\\ \text{in vial}\end{array}\right\rangle\left|\begin{array}{c}\text{cat still}\\ \text{alive}\end{array}\right\rangle\right)
\end{aligned}$$

Further, our prequantum theories give us adequate understanding of the processes

$$\begin{aligned}
&\left|\begin{array}{c}\text{atom}\\ \text{undecayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter}\\ \text{untriggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison}\\ \text{in vial}\end{array}\right\rangle\left|\begin{array}{c}\text{cat}\\ \text{alive}\end{array}\right\rangle \longrightarrow\\
&\qquad\left|\begin{array}{c}\text{atom}\\ \text{undecayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter still}\\ \text{not triggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison still}\\ \text{in vial}\end{array}\right\rangle\left|\begin{array}{c}\text{cat still}\\ \text{alive}\end{array}\right\rangle
\end{aligned}$$

and

$$\begin{aligned}
&\left|\begin{array}{c}\text{atom}\\ \text{decayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter}\\ \text{untriggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison}\\ \text{in vial}\end{array}\right\rangle\left|\begin{array}{c}\text{cat}\\ \text{alive}\end{array}\right\rangle \longrightarrow\\
&\qquad\left|\begin{array}{c}\text{atom}\\ \text{decayed}\end{array}\right\rangle\left|\begin{array}{c}\text{counter}\\ \text{triggered}\end{array}\right\rangle\left|\begin{array}{c}\text{poison}\\ \text{released}\end{array}\right\rangle\left|\begin{array}{c}\text{cat}\\ \text{dead}\end{array}\right\rangle .
\end{aligned}$$

It is then tempting to interpret (1) as describing the evolution of two (or two sets of identical) parallel worlds, both primarily evolving in a classical way but sometimes interfering with one another. This temptation is strengthened when we put ourselves into the formalism: a human observer of the cat becomes entangled with it, and the universal state is a superposition of definite observations.

[10] The reason for distinguishing "counter untriggered" from "counter still not triggered" (and similarly for the cat and the vial) is somewhat pedantic: being macroscopic systems, these objects will inevitably evolve to some degree over any time interval (the cat will metabolise food, for instance). In fact, strictly speaking it is unlikely to be correct even to regard these macroscopic objects as pure states: realistically they will be significantly entangled both with the environment and with each other.

As was mentioned in the Introduction, there are serious problems with taking this parallel-worlds viewpoint literally; let us see how these play out in more detail. The greatest difficulty is probably the need to specify which states describe single worlds and which describe superpositions of them. The only way to answer this question — if we take the individual worlds as ontologically primary — is to specify a 'preferred basis' ($\{|i\rangle\}$, say) for the universal Hilbert space. Then any state $|\psi\rangle$ can be expressed in terms of this basis:

$$|\psi\rangle = \sum_i \alpha_i |i\rangle$$

and interpreted as a collection of parallel worlds (indexed by i) each of weight α_i.

Four problems present themselves. Firstly what do we mean by the phrase "each of weight α_i"? It is not sufficient to take it as saying that the fraction of world i in the collective is $|\alpha_i|^2$, since this discards the relative phases of the component worlds, which play an important rôle in the theory.

Secondly, how is this $\{|i\rangle\}$ basis selected? To justify the reasoning which led us to Everett in the first place, it had better be a basis whose members are not states describing macroscopic objects as delocalised, but this restriction is far from enough to specify it. Making some choice by fiat is hard to motivate: the configuration-space basis is one obvious choice, but violates relativistic covariance by requiring a preferred reference frame.[11]

The third and fourth problems concern the fact that these 'worlds' do not have all of the properties of worlds in the prequantum sense. Our third problem is that they are instantaneous: we may decompose the universal state into worlds at some given instant of time, but we cannot track the individual worlds when we evolve the state forward in time in any satisfactory way. All we can say is that world 14 increases in weight, while worlds 15 and 16 decrease and world 17 changes phase, etc. but we have no classical notion of one set of definite properties passing into another. (To see this, suppose that the state $|\psi\rangle$, above, has evolved over time t into a new state

$$|\psi'\rangle = \sum_i \alpha'_i |i\rangle\,;$$

if each $|i\rangle$ represents a world which evolves over time then we need some notion that world $|i\rangle$ has evolved over time t into some other world $|i'\rangle$, but the formalism of quantum mechanics will not give us this — it just tracks the change in weights of the time-invariant states $|i\rangle$. To recover such a notion would (as

[11]There's also a slightly subtler problem with this choice: generalised to field theory, the "position" basis becomes the basis of definite field configurations. But this basis fails to describe macroscopic objects properly, since they are made up of *particles*, which are specific superpositions of field-configuration eigenstates. Choosing a basis of definite particle locations in QFT is also problematic (Saunders 1992), because (i) the concept of location is only approximate for relativistic particles; (ii) given mass renormalisation, which states describe particles depends on the energy at which we observe; (iii) in general spacetimes, the particle basis cannot be exactly defined even for free quantum fields.

Barrett (1999) has argued) require us to supplement unitary quantum mechanics with not only a preferred basis, but also some auxiliary dynamical rules — but by this stage we are far from the Everett interpretation *per se*.)

The fourth problem is that our thoughts do not supervene on individual worlds any more than they do on individual moments of time: whether we identify the objects of supervenience as states or processes of the brain, the projector onto (neurologically) functionally identical brain states will be very high-dimensional — for how can it matter to neurology whether a certain electron in my frontal lobe is displaced by 10^{-15} metres? For all the richness of human experience, there surely cannot be enough distinct conscious states to correspond one-to-one with the number of orthogonal states corresponding to a working brain; I shall argue that (at least if we are functionalists) this forces us to identify thoughts as supervening on facts about certain regions in large numbers of similar but non-identical worlds.[12]

Why are we forced to such an identification? If the worlds did not interact at all then we could possibly avoid it: it is at least defensible to suppose that there exist many distinct non-interacting worlds containing many distinct but functionally identical copies of me.[13] The problem is that for any simple choice of the preferred basis, interference between different terms in the basis is vital to the dynamics of the processes in my brain (or indeed any macroscopic object). If we choose the position basis as preferred, for instance, we might decompose a certain brain state (describing my brain in a certain definite configuration) into position eigenstates and suppose that each is a separate conscious entity. But if one such eigenstate were isolated from the rest (if an external observer measured the position of all brain constituents, for instance) then that state would evolve rapidly into organic soup and certainly would cease to be a functioning brain.[14] Hence if mental facts are supervenient on facts about brain configurations in single worlds then each of us remains a thinking being only because of constant interference from the particles comprising the brains of countless neurologically identical parallel-world copies of ourselves. This is possibly not an untenable view (so long as we hold on to the idea of worlds as fundamental) but to me it seems an unattractive one. (The only way to avoid this problem would be to take the preferred basis as a decoherence basis (see section 5) for the brain, in which case interference between the different terms will be negligible — but this would require the use of detailed biochemical and neurophysiological criteria to specify what is supposed to be a basis which is preferred at the level of basic ontology, and even then the interference terms will not be completely eliminated.)

[12] Such an identification has previously been argued for by Donald (1997):

> In as far as there are such sets of *identical* experiences, I would associate them with the *same* mind. [Donald's emphasis]

[13] In fact, this viewpoint is held by modal realists such as Lewis (1986).

[14] In case this is not obvious, consider the bonds holding together the molecules of the brain: each is made up of various electrons in highly delocalised states. A localised electron, then, is conversely in a superposition of binding, and not binding, a given pair of atoms. Breaking a significant fraction of the bonds in every complex molecule in the brain would probably not be the only deleterious effect of a position measurement, but it will do for a start.

This is to say nothing of what lies outside my skull — but if some object many light years away is in a superposition of two preferred-basis states then taking the worlds literally implies that there are two sets of parallel worlds with the copies of my brain in one set identical to the copies in the other set. In this situation are there two identical versions of me, or should I regard my thoughts as supervening on both simultaneously? There will obviously be no detectable difference between these two viewpoints, but in the former the state of the faraway object will be determinate but unknown and in the latter there will be no fact of the matter about its state.

Generally speaking, modern versions of Everett get round this problem as follows: they abandon the idea of specifying a preferred basis explicitly, and then recover it — in an approximate or pragmatic sense — from the state and the dynamics by one consideration or another. The usual consideration — and the one which will be adopted here — is decoherence theory (Zurek 1991; Saunders 1995), which tells us that subsystems of the world will have 'states' (i. e. reduced density operators) diagonalized with respect to a certain basis, and that attempts to prepare these subsystems in states not belonging to this basis will be virtually impossible. The basis thus determined, however, is *approximate*: many choices of basis will satisfy the decoherence property with (for instance) the choice of basis for the spin of a given electron being irrelevant unless that electron's spin is entangled with some macroscopic system (such as a Stern-Gerlach apparatus).[15]

But in rejecting any idea of worlds as ontologically primary, we have returned to the problem at the beginning of this section: if the state is fundamental, how are we to understand it? Hopefully the parallels with spacetime are clear, and their implications equally so: we are to understand the universal state as

> an entity with such-and-such mathematical description *which can be thought of as a collection of instantaneous worlds together with their Hilbert-space amplitudes, no matter that it may be decomposed into such a collection in many different ways.*

As with the instants of spacetime, the point is not that we are directly presented with worlds of definite particle position through our senses (we aren't, or not exactly), nor that we ourselves live in a single such world (we don't) but that we have a conceptual grasp on the idea of such worlds. We can regard the universal state as being made up of worlds and their amplitudes in the same way that we can regard spacetime as made up of instants and their metric relations: neither description really does justice to the symmetries of the entity being described, but both give enough data completely to specify the entity and both give us a conceptual grasp of what this entity is.

In this way, the analogy with spacetime and the many-instants view allows supporters of Everett to answer the charge that, without some explicitly pre-

[15]Furthermore, it is possible that it is to some extent anthropocentric: it is impossible for functional systems like us to prepare the subsystems in states not belonging to the basis, but it remains unproven that other, wildly different systems would face the same restriction. I am grateful to Simon Saunders for this point.

ferred basis, their world-view simply does not make sense (as distinct from the question of whether, for instance, it is empirically valid). For the idea of a universal state regarded as ontologically prior to worlds makes no more and no less sense than the idea of spacetime regarded as ontologically prior to instants of time, and if we insist that the Everett interpretation requires a preferred basis to make sense then we ought to insist also that relativity requires a preferred foliation.[16]

5 Consistent Histories and the Persistence of Worlds

Let us examine the concept of this "effectively preferred basis" more carefully. A basis $\{|i\rangle\}$ can equally be expressed as a partition of unity

$$\widehat{1} = \sum_i \widehat{P}_i$$

where $\widehat{P}_i$ projects onto the one-dimensional subspace spanned by $|i\rangle$. We can characterise such partitions as being a collection of projectors which

1. sum to unity;
2. are mutually orthogonal, $\widehat{P}_i \widehat{P}_j = 0$ for $i \neq j$; and
3. project onto one-dimensional subspaces.

We can then understand the idea of an "approximate basis" mentioned above by dropping the third of these requirements. Decoherence specifies such a collection of projectors, and it is an approximate basis in the sense that any exact basis which (regarded as a resolution of unity) is a fine-graining[17] of this collection, will consist of states which are stable against the decoherence process. By this, we mean that even if at some time a state (expressed as a density operator) is not block-diagonalized by the decoherence projectors, the off-block-diagonal matrix elements of the operator will decay on a vanishingly short timescale compared to the timescale on which the block-diagonal matrix elements evolve. The information encoded in the off-diagonal states disappears into entanglement between states of macroscopic systems and is effectively inaccessible;

As is well known, the consistent histories formalism (Griffiths 1984; Gell-Mann and Hartle 1990; see Kent (1998) for a recent review) abstracts this requirement and frees it from direct reference to subsystems and to the decoherence process. A consistent history space is, roughly, a collection of coarse-grained bases (with the collection indexed by time, so that there is one basis for each moment of time[18]). The consistency condition ensures that we can apply

[16] It might be argued that Barbour (1999) requires both: his metaphysics (for reasons connected to spacetime relationism and quantum gravity) includes both a preferred foliation and a choice of configuration space as the preferred basis.

[17] If we have two resolutions of the identity $\{\widehat{Q}_i\}$ and $\{\widehat{P}_j\}$, the latter is a fine-graining of the former if each $\widehat{Q}_i$ is a sum of some of the $\widehat{P}_j$.

[18] Actually, in practice we often discretise time rather than using the continuum.

classical probability to the events described by the projectors in these bases (in the sense that each history can be assigned a probability, and if history A is a coarse-graining of histories $B_1, \ldots B_n$ then the probability of history A is the sum of the probabilities of its component histories).

The problem given rise to by this abstraction is that there exist many choices of consistent history space, but if we follow Everett and keep the state as fundamental there is no problem. Just as our choice of world-decomposition (i. e. fine-grained basis) is made for ease of description rather than more fundamental reasons, so our choice of history space is just made so as to give a convenient description of the quantum universe. In fact there will be a subset of history spaces which are *much* more convenient: we are information-processing systems, and it can be shown that any such system picks out a consistent history space (Saunders 1993). (Reverting to the subsystem description given earlier, the point is (in part) that such a system needs to store memories and if it chooses an encoding of memories into states which are not diagonal in the decoherence basis, they will not last long (Halliwell 1993; Zurek 1991).) So for describing events in our vicinity, at least, there is an overwhelmingly preferred choice. As with the pragmatically preferred reference frames of relativistic spacetimes, the preference is only approximate and really only extends to our spatial vicinity: if we wish to pick a truly global, fine-grained basis then there will be considerable arbitrariness.

Let us now reconsider the four problems with many-worlds identified in the previous section. Of these, three have close analogues in the many-instants description of spacetime: in describing reality there is neither an uniquely best set of instants, nor of worlds; we cannot recover spacetime from its instants without their temporal relations any more than we can recover the universal state from the worlds without their amplitudes; and mental facts supervene neither on facts about a single instant nor about a single world. Hence if we accept the many-instants description of spacetime regardless, none of these problems prevent us taking the Everett interpretation seriously, nor explicating it (with appropriate care) in terms of worlds.

However, the third problem identified in section 4 remains: the (fine-grained) worlds in a superposition are instantaneous, and we have not yet addressed the question of how (or whether) we can track a given world from moment to moment. But this situation should be familiar to us from spacetime physics: there, too, the notion of a persisting object is not directly present in the formalism. However, the notion can emerge given a sufficiently ordered spacetime: we may be able to pick out 'world-tubes' of fairly stable matter configurations, and declare different three-dimensional slices of that tube to be the same object at different times. This concept may not be definable with arbitrary precision (how much is an object allowed to change before ceasing to count as *the same* object?) and will not be applicable if the spacetime is not ordered enough for persistent world-tubes to exist; nonetheless it is useful.

We can use the same pragmatic concept of continuance to make sense (in certain circumstances) of the idea of identifying worlds across time. In neutron interferometry, for instance, we have a neutron in a linear superposition of two spatially separated wave-packets. Until they are brought back together again these packets do not overlap, so the evolution of each packet in position space can be treated independently, without allowing for interference between the two, and so we can reasonably describe the neutron as being in a superposition of two histories, each fairly well-defined. This makes it reasonable to speak of two persisting worlds, each describing one neutron. (See Vaidman (1998) for a detailed discussion of neutron interferometry in the Everett interpretation, from a related viewpoint.) Of course it is vital to remember that this language does not describe anything fundamental in the theory, and that it will fail in certain circumstances (in this example, it will fail when the two neutron beams are brought back together again, so that interference occurs between the two worlds) or when looked at too closely (in the example, we may speak of two persisting worlds, but the worlds are fairly coarse-grained). Nonetheless the concept of persisting worlds may be useful in certain explanatory contexts (rather like the concept of 'living creature' in biology, or 'star' in astrophysics, neither of which have an unambiguous definition or are written directly into the formalism[19]).

In this context we see the consistent-histories formalism from another viewpoint: consistency gives a criterion for (coarse-grained) worlds for which there exists a fairly robust notion of persistence. By 'fairly robust' I mean that although we may need to speak of worlds splitting[20] and (in thermodynamically implausible circumstances) recombining, we do not have to deal with the interference between worlds that generally makes persistence meaningless.

6 The Status of the Observer

As was stated in the previous section, if we have a consistent-history basis (at time t) $\{\widehat{Q}_i(t)\}$ then it is constrained by the requirement that our brains (being information-processors) carry out this processing in such a basis. Then I can identify a set of projectors $\{\widehat{R}_j\}$ which are a coarse-graining of $\{\widehat{Q}_i(t)\}$ and which project onto functionally distinct states of my brain.[21] This set of projectors is not a resolution of the identity, since that would be to assert that my brain exists with probability one; furthermore it is *very* coarse-grained in comparison with $\{\widehat{Q}_i(t)\}$. After all, specifying a set of projectors onto states of my brain does not involve giving any information about the rest of the universe.

We now have a three-level description of the quantum state.

At the lowest level, we have a description in terms of **Everett worlds**, i. e. by means of a fine-grained basis. It is at this level that we are giving a *complete* description of the state; nonetheless these worlds are instantaneous

[19]See Deutsch (1997) and Dennett (1991a) for further discussion of this issue.

[20]The idea of worlds splitting will obviously lead to ideas of a one-many criterion for identity over time, both for the worlds and for the macroscopic objects within them. Though there are objections to such criteria, especially as regards the interpretation of probability (see, e.g., Barrett (1999)), they have been discussed in detail elsewhere (in particular by Saunders (1998) and Tappenden (2000b)) and will not be dealt with further here.

[21]If we are interested in questions of supervenience, it may well be that mental facts are supervenient upon facts about processes, not states. In this case we would need to consider sequences of coarse-grainings.

entities with no traceable histories, and there is a high degree of arbitrariness about the choice of basis.[22]

The arbitrariness is not complete, however, since to be practically useful this basis must be a fine-graining of some **decoherence basis**, which will in turn be given by some choice of consistent history space. At this level of the description we can talk usefully of histories, describing a branching (and occasionally recombining, in principle but not in practice) set of worlds. This is the level at which we obtain a useful classical limit. However a description of the world at this level is not complete, since the coarse-grained nature of the decoherence basis means that information is lost.

In turn, the choice of consistent history space is motivated by anthropic considerations: we are information-processing systems and as such must be embedded in some consistent history space in which this information processing takes place (Halliwell 1993; Saunders 1993). As such a useful choice of consistent history space must be a fine-graining (at each moment of time, or for each short space of time if we wish to think of thoughts supervening on processes rather than states) of the **basis of functionally distinct brain states**. At this level our description of the state is highly incomplete. However, at least from an epistemological view point if the universal state is $|\psi\rangle$ and I am in brain state j (with $\widehat{R}_j$ being the projector onto this state) then I should regard the state of the Universe relative to me as given (up to normalisation) by $\widehat{R}_j\,|\psi\rangle$.

Note the contrast between the way in which observers are recovered from unitary quantum mechanics in this sort of approach (and in the closely related approaches of, e. g., Saunders (1996a), Zurek (1991), Gell-Mann and Hartle (1990)), compared to the approach taken by many-minds theorists like Lockwood (1989, 1996a) or Donald (1995, 1997, 1999).[23] In the latter approach, the projectors $\widehat{R}_j$ are introduced from the outset, and the interpretation is constructed in terms of them; in the former, the job of physics is taken to be the recovery of a quasiclassical domain and the problem of consciousness is then handed over to other disciplines.

To some extent, reasons for pursuing one or other scheme come down to differences of opinion about how successful certain scientific or philosophical programs will be, so that decoherence-based approaches depend on the success of the decoherence program to succeed formally in recovering an approximately classical physics, while many-minds theorists can ignore this program but are instead committed to certain quite strong physical or philosophical views about consciousness (so Lockwood's approach requires there to be a fact of the matter as to which physical systems possess consciousness bases,[24] while Donald (1995)

[22] As Mermin (1998) has shown, we can equally well give a description at this level in terms of local events and the correlations between them: given any resolution of a system into subsystems, knowing the correlations between observables in the subsystems gives us sufficient information to recover the state.

[23] Albert and Loewer (1988), though they also advocate a many-minds theory, differ from Lockwood and Donald in that they add explicit extra structure in the form of continuing minds.

[24] cf. Lockwood (1996b):

has constructed explicit models of observers and holds that this is a necessary part of interpreting Everett[25]).

It is obviously far too early to say which way these longstanding controversies will resolve themselves; however, section 2 seems to give some support to the decoherence approach: whether or not mind turns out to be essential in interpreting quantum mechanics, we will have to recover a quasiclassical domain *anyway* to make contact with existing theories and so interpret quantum theory as a theory of physics. If there are no theory-neutral observations, then saving the observed phenomena requires us to save (something approximately isomorphic to) the theory.

I mention in passing an ambiguity about our choice of $\widehat{R}_j$. These projectors are designed to pick out *my* brain, but how is that to be done? Presumably my thoughts can be instantiated in many different physical media; they can certainly exist in many different spatial locations and be composed of different atoms, so requiring them to project onto a given subsystem's Hilbert space isn't really plausible. Presumably I have to use some intrinsic structural features, but there seems to be an arbitrariness as to how great a personality-shift can be tolerated before a given projector should actually be counted as projecting onto *someone else's* brain state. This is, of course, the old problem of personal identity in quantum-mechanical guise.

7 Covariant Quantum Mechanics

So far, we have constructed an analogy between spacetime, and the quantum state *at a given time*. However, talking of worlds at a given time obviously implies a choice of reference frame and a breaking of covariance, whereas one of the major motivations for adopting the Everett approach is its claim to provide a quantum mechanics which is compatible with relativity. Accordingly, we now turn to the problem of understanding relativistic quantum theory.

The mathematical structure given for quantum field theory at the beginning of section 4 is fully covariant, since it is written in the Heisenberg formalism so that its states are not dependent on time. Hence we have an uninterpreted

> I regard it as a straightforward matter of fact whether cats or (a more seriously contentious issue) prawns are sentient. Thus it's not, as I see it, a matter of how we define our terms, but whether there's a 'what it's like to be' a cat or a prawn [...]

If, in fact, the correct theory of consciousness were to be a much less all-or-nothing affair (e.g. Dennett (1996):

> the features [of animal consciousness] we have examined seem to make their appearance not just gradually but in an unsynchronised, inconsistent and patchy fashion, both in evolutionary history and in the development of individual organisms.)

it would presumably have a significant impact on Lockwood's theory.

[25]cf. Donald (1997):

> [...] the proper task of an analysis of quantum theory is to give an exact definition of the possible physical manifestations of an observer.

object — the state ρ — which has the properties we require. To describe this object we can combine the many-worlds and many-instants descriptions:

> The universal state is an object with such-and-such mathematical description *which can be described as a collection of sets of worlds, together with the temporal metric relations between sets and the amplitudes of individual worlds within each set, no matter that the universal state can be decomposed into such a collection in many different ways.*

There is an attempt to illustrate this viewpoint on page 277 of Deutsch (1997), although this illustration gives the impression that individual worlds have histories (i. e. that it is possible to say that *this* world in *this* instantaneous collection of worlds is the future version of *that* world in *that* collection). This is consistent with Deutsch's explicit introduction of a preferred basis in Deutsch (1985), and with the 'continuing minds' of Albert and Loewer (1988) (in both of which there exist persisting objects — be they minds or worlds — which are stochastically partitioned as the wave-function evolves), but contrasts with the view presented here, in which persisting worlds are an approximate and derived notion. (Obviously we could justify such a description in this framework by working at the level of consistent histories, but then we cease to have a complete description of the full instantaneous collection.) Deutsch also makes the speculative suggestion (inspired by canonical quantum gravity and by the work of Page and Wootters (1983); see also Barbour (1999)) that this two-axis collection of worlds can be collapsed into a single collection — in other words, that different instants can be understood as different Everett worlds — but this remains controversial, and lies beyond the scope of this paper.

It must be admitted that our description of the universal state, given as it is in terms of global instants of time and worlds defined at those instants, is somewhat unaesthetic. In the next two sections I shall develop a somewhat modified version which better reflects relativistic covariance, and then consider an important strength of our existing description.

8 Spacetime as the Set of Events

One partial method of emancipating special-relativistic physics from the language of space-at-an-instant is to see it as the set of events and the spatiotemporal relations between them.[26] This approach deserves mention here, as it succeeds partially in conveying a direct intuition of relativistic spacetime but is still reliant in many ways on our prerelativistic intuitions.

The easiest way to see this is to look more closely at the concept of 'event' — by which we mean something that occurs in a given place, at a given time (an explosion, say, or a flash of light, or the coincidence of hands on a clock). But

[26] Specifying these spatiotemporal relations is crucial: without them we have no way of distinguishing numerically distinct, but identical, events (such as two identical flashes of light in different places). In this (rather Leibnizian) sense an event is really a global notion — to specify it, we need to specify its relations with all other events.

this description already makes intensive use of prerelativistic concepts, both in its direct use of space and time language and in its reference to physical objects. We could of course define events as crossing-points of world-lines, but now we have either to explain what world-lines represent (surely impossible without our prerelativistic language) or to return to uninterpreted mathematics ('a world-line is a smooth map of the real line into the four-manifold such that the metric distance of any two points in the image is positive', etc.).[27]

However, one thing this discussion reminds us of is that our use of prerelativistic intuitions does not require a *global* notion of space. We cannot say what we mean by an event without any reference to the space/time split but we can describe it with reference to space and time only in its vicinity.

As such we may understand relativistic spacetime by

1. breaking it up into small regions;
2. describing each region in terms of a (highly non-unique) space-time split — i. e. a local version of many-instants;
3. giving the spatiotemporal relations between the regions.

The pure-event viewpoint can be understood as a limiting case of this interpretation scheme, although in the limit we once again lose our ability to understand the theory (as has been argued for in section 3).

This method of understanding relativistic spacetime has a number of advantages over the use of globally defined instants: in particular it brings out the arbitrariness of our choice of spatial slice at large distances and the irrelevance of that choice in describing local phenomena, and it makes it clear that there is no fact of the matter about what is happening *now* in distant regions of space. However, it is not really a more "right" description of spacetime — the best which we might be able to say about it is that it is a more useful description of an entity whose perfect description as a physical system lies (at least for the moment) beyond our ability to comprehend directly.

The set-of-events description of spacetime can be copied in the quantum case. An (instantaneous) 'event' in quantum theory is specified by a (Heisenberg-picture) projector (onto states for which, at the time at which the event is supposed to occur, it does indeed occur with certainty). If events A, B are described by Heisenberg projectors $\widehat{P}_A, \widehat{P}_B$ respectively, then the probability of event A occurring given that event B occurs is

$$\frac{\mathsf{Tr}(\widehat{P}_A \widehat{P}_B \rho \widehat{P}_B \widehat{P}_A)}{\mathsf{Tr}(\widehat{P}_B \rho \widehat{P}_B)}.$$

[27] The subtleties of defining spatially and temporally extended events were recognised by Einstein (1905), whose original discussion of special relativity refers in passing to

> the imprecision that is inherent in the concept of simultaneity of two events taking place at (approximately) the same location and that must be surmounted by an abstraction.

The issue is discussed further by Saunders (1997).

We are now in a position to regard the quantum state as (being described by) the collection of events together with the spatiotemporal and probabilistic relationships between them.

This relational view of Everett has been recently championed by Saunders (1993, 1995, 1996b, 1997, 1998). As with the classical description of spacetime as events and spatiotemporal relations between them, it is in many ways less arbitrary than using global 'instants' and 'worlds' but still relies on a (classical, pre-relativistic) concept of what an event is. Essentially the concept which we import from our previous theories is that of a 'local world', i. e. a small spatial region, at a single time, in which certain observables have definite values (note the explicit need to choose a local simultaneity convention, for self-adjoint projectors can project onto spatially extended events but not onto temporally extended ones).[28]

9 Telling a Dynamical Story

The 'local' description given in the previous section would appear to have significant advantages over the many-instants approach, in that it better captures the notions of covariance and locality which are central to relativity. Should we then abandon the many-instants description of section 7 in favour of the local description?

We can see why such abandonment would be unwise when we consider the issue of dynamics. We need the notion of a global foliation of spacetime (whether we are doing classical or quantum physics) if we wish to discuss spacetime physics as a dynamical system, and in particular if we wish to discuss issues of determinism. A globally specified instant, together with the (classical or Schrödinger) state at that instant, incorporates an important feature of our prerelativistic theories which is less transparent in the local description: namely that at any given moment of time we can specify the contents of space everywhere without having to worry that one region has causal influence upon another, but that having done so, at later moments of time the contents of space will be determined by the dynamics.

In discussing dynamics in relativity, we are again required to behave in this way: we need to specify the matter distribution and its rate of change across a spacelike surface (i. e. a choice of instant) which extends across the whole region of interest (in principle across the whole universe) and then evolve this spacelike surface forward in time (Wald 1984). There are, of course, a huge number of different ways in which we may specify this time evolution, each generating a different choice of spacetime foliation and hence a different 'many-instants' description of spacetime (this is the 'many-fingered time' of Misner, Thorne, and Wheeler (1973)).

The covariance of relativity theory manifests itself in our ability to describe

[28] Recent work by Deutsch and Hayden (1999) on the Heisenberg picture in quantum computation has given rise to some suggestions of another entirely local — and resolutely nonclassical — way to understand unitary quantum mechanics, but this work remains in its infancy.

a dynamical process equally validly by many different foliations. However, only in the most stylised of dynamical processes in classical or quantum relativity can we give such a description in a manifestly covariant way which makes no use of a foliation[29]: the very concepts of 'process' and 'dynamics' depend upon the notion of something taking place as time changes, and so cannot be given without reference to time.

As such, if we wish to describe causal or dynamical processes in a given region, we must make a space/time split which is global at least over that region. So if we were describing the evolution of life on earth, for instance, we may be able to restrict our slices to some neighbourhood of the solar system — and indeed can get quite far by restricting them to a neighbourhood of the Earth's surface. Only when we do cosmology do we need to foliate the entire spacetime.

How much freedom do we have in choosing our foliation, and our set of worlds? For a large class of dynamical processes (essentially all those taking place in our local vicinity, at macroscopic scales and low velocities) there is (locally) a fairly well defined 'best choice' of foliation and of basis common to all members of the class: the foliation picked out by Lorentz reference frames co-moving with the Earth, and the basis picked out by decoherence. Many other processes, while not sharing these choices, still have an obviously 'best' choice: for classical relativity, consider the atmospheric physics of a star moving at relativistic speeds relative to us; in quantum physics, consider neutron interferometry with its fairly well-defined separate histories. Other processes again cannot be properly understood until we have seen them described in several distinct bases or foliations: consider:

- colliding shock-waves, where we need to consider both the centre-of-mass viewpoint and that of an observer moving with the shock;
- gravitational collapse (Misner, Thorne, and Wheeler 1973), where both the description from a surface co-moving with the collapsing star and the description of a faraway observer give important insights;
- the two-slit experiment, which notoriously cannot fully be understood from either a particle or a wave viewpoint;
- certain algorithms in quantum computation (Deutsch, Ekert, and Lupacchini 1999) (indeed, from a certain viewpoint every algorithm in quantum computation, since quantum computers outpace classical computers partly through not being forced to work in a fixed basis).

The moral is that we should not regard 'local' and 'many-instants' descriptions, whether of classical or quantum physics, as rivals: rather, they illuminate dif-

[29]Contrast the two-body collisions of otherwise free particles (adequately described in terms of world-lines) with computer models of supernovae (Arnett 1996), or the (admittedly complex) scattering theory of asymptotically free relativistic quantum particles with the far more intractable lattice simulations used to analyse non-perturbative quantum field theory (see Kaku (1993) for an introduction).

ferent aspects of the (classical or quantum) universe, and which one is used will depend on the specific situation.

10 Summary: a Point-by-Point Comparison

The rich analogies between spacetime and the Everett interpretation stand out clearly in the following table:

Relativity	Everett
The Universe may be fully described as a collection of instants and their temporal relations. Such a description is arbitrary and fails to show the full structure of spacetime.	The Universe may be fully described as a collection of (fine-grained) worlds, and their weights. Such a description is arbitrary and fails to show the full structure of the multiverse.
The instants are theoretical constructs and our thoughts do not supervene on individual instants. Nonetheless our familiarity with our pre-relativistic theories gives us a conceptual grasp on the idea of 'now' and an instant,which enables us to understand what is described by spacetime.	The worlds are theoretical constructs and our thoughts do not supervene on individual worlds. Nonetheless our familiarity with our pre-quantum theories gives us a conceptual grasp on the idea of a world which enables us to understand what is described by quantum theory.
Specifying only the collection of instants without giving the temporal relations between instants is insufficient to tell us the spacetime.[30] We know mathematically how to handle spatiotemporal relations, and we have a practical understanding of temporal duration and flow, but these concepts are philosophically problematic. We may speak of 'moments of time' and the number of moments of time ('the next moment', etc.) but this is just a metaphor for temporal duration, and cannot be interpreted literally.	Specifying only the collection of worlds without giving the amplitudes of each world is insufficient to tell us the state. We know mathematically how to handle amplitudes, and we have a practical understanding of probability, but these concepts are philosophically problematic. We may speak of 'number of worlds' but this is just a metaphor for the weight of a given world, and cannot be interpreted literally.

[30]See Barbour and Bertotti (1977, 1982) and Barbour (1999), however, for further discussion of this point from a Machian perspective.

Relativity	Everett
The theory is specified in a way which makes no direct reference to observers, but to understand why the Universe seems as it does to *us* we need to address certain observer-related problems — specifically we need to understand our perceptions of the passage of time.[31]	The theory is specified in a way which makes no direct reference to observers, but to understand why the Universe seems as it does to *us* we need to address certain observer-related problems — specifically we need to understand our perceptions of probability.
In everyday circumstances, there exists an (approximate) natural choice of choice of spacetime foliation.[31] The details of this foliation are arbitrary, both when examined closely and with respect to spatially remote areas.[31]	In everyday circumstances, there exists an (approximate) natural choice of basis. The details of this basis are arbitrary, both when examined closely and with respect to spatially remote areas.
In describing the dynamics of a system it is generally necessary for us to give our description in terms of some choice of foliation (i. e. instants). For some processes (such as a boardroom meeting or the dynamics of the Solar System) there is an approximately-defined 'best' choice of foliation; for others (such as gravitational collapse) different choices may give different insights into the process.	In describing the dynamics of a system it is generally necessary for us to give our description in terms of some choice of basis (i. e. worlds). For some processes (such as neutron interferometry, or a Schrödinger's Cat experiment, or (possibly) making sense of counterfactual reasoning) there is an approximately defined 'best' choice of worlds; for others (such as a quantum computation) different choices may give different insights into the process.
There is no fundamental notion of an object's persistence through time; all we have are structural similarities between regions of space at different times.[32] However, in certain situations it may be possible to recover a (pragmatic, approximately defined) notion of persistence of objects from these structural features.[32]	There is no fundamental notion of a world's persistence through time; all we have are structural similarities between parts of the state at different times. However, in certain situations it may be possible to recover a (pragmatic, approximately defined) notion of persistence of worlds from these structural features.

[31] These matters are discussed further by Stein (1991).

[32] It seems likely that these notions apply to our own concept of personal identity as much as to physical objects; see Parfit (1984) for a defence of this.

Relativity	Everett
We may describe the classical spacetime as a network of events together with the spatiotemporal relations between the events. Whether we describe spacetime in terms of many global instants or in terms of more localised events, we make essential use of our intuitive concept of a local spatial region — not in the mathematical formulation of the theory but in getting a conceptual grasp of it and understanding our own place in it.	We may describe the quantum spacetime as a network of events together with the spatiotemporal and probabilistic relations between the events. Whether we describe the quantum universe in terms of many global worlds or in terms of more localised events, we make essential use of our intuitive concept of a local and value-definite (for some observables) spatial region — not in the mathematical formulation of the theory but in getting a conceptual grasp of it and understanding our own place in it.

11 Conclusion

Are the worlds real? Yes, in the sense that instants of time are real: they may not be present directly in the formalism, but unless we introduce the concept we may struggle to understand what the formalism is telling us, and without the concept it may be impossible to capture important (causal/deterministic) properties of the world. Further, a description in terms of worlds (or instants of time) is not *incomplete*, for we can certainly recover the universal state from it; our only complaint with it is that it is somewhat arbitrary.

Are consistent histories, and worlds which persist over time, real? Yes, in the sense that rivers, or animals, or persisting objects, are real: like worlds or instants they are not directly present in the formalism, and unlike worlds or instants they only approximately definable, but that is no reason why they should not be seen as legitimate entities or used in our explanations (any more than we should expect to be able to describe zoology in any useful or explanatory way using only the language of quantum field theory).

We are undoubtedly more at home with Minkowski spacetime than with the universal state. Partly this may be because we have worked with the concept in physics for rather longer, but more importantly we have long been used to the idea that multiple times exist (in some sense) — the innovation in relativity theory is the unification of these instants into a whole, and the identification of the instants as secondary concepts. Everett asks us to take both steps at once: to accept that there exist many worlds,[33] and then to fuse them together into a whole and accept that the worlds are only secondary. Clearly this is a significantly larger conceptual jump; still, if we are prepared to accept the existence of many worlds and if we are happy with the step from many times to

[33]Of course, this is not a wholly new concept: whatever we may think of the metaphysical status of possible worlds, we routinely describe them when we use counterfactuals.

spacetime, there seems no reason to avoid a similar step in the case of quantum theory.

Acknowledgements—For extensive discussions, suggestions and comments on this work, I am very grateful to Hannah Barlow, Katherine Brading, Harvey Brown, Jeremy Butterfield, Chris Fuchs, Clare Horsman, Lev Vaidman, and especially to Simon Saunders, whose work on Everett has been a major influence on this paper. I am also grateful to David Deutsch for many discussions of the Everett interpretation.

References

Albert, D. and B. Loewer (1988). Interpreting the many worlds interpretation. *Synthese 77*, 195–213.

Arnett, W. D. (1996). *Supernovae and nucleosyntheses: an investigation of the history of matter, from the Big Bang to the present.* Princeton: Princeton University Press.

Barbour, J. (1999). *The end of time.* London: Weidenfeld and Nicholson.

Barbour, J. and B. Bertotti (1977). Gravity and Inertia in a Machian framework. *Nuovo Cimento 38B*, 1–27.

Barbour, J. and B. Bertotti (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society of London 382*, 295–306.

Barrett, J. (1999). *The quantum mechanics of minds and worlds.* Oxford: Oxford University Press.

Bohm, D. (1952). A suggested interpretation of quantum theory in terms of "hidden" variables. *Physical Review 85*, 166–193.

Butterfield, J. (1996). Whither the minds? *British Journal for the Philosophy of Science 47*, 200–221.

Dennett, D. (1991a). Real patterns. *Journal of Philosophy 87*, 27–51. Reprinted in *Brainchildren*, D. Dennett, Penguin, 1998, pp. 95–120.

Dennett, D. (1996). *Kinds of Minds: towards an understanding of consciousness.* London: Phoenix.

Dennett, D. C. (1991b). *Consciousness Explained.* London: Penguin.

Deutsch, D. (1985). Quantum Theory as a Universal Physical Theory. *International Journal of Theoretical Physics 24*(1), 1–41.

Deutsch, D. (1997). *The Fabric of Reality.* London: Penguin.

Deutsch, D., A. Ekert, and R. Lupacchini (1999). Machines, Logic and Quantum Physics. Available at http://www.arxiv.org/abs/math.HO/9911150.

Deutsch, D. and P. Hayden (1999). Information Flow in Entangled Quantum Systems. Available at http://www.arxiv.org/abs/quant-ph/9906007.

DeWitt, B. (1970). Quantum Mechanics and Reality. *Physics Today 23*(9), 30–35. Reprinted in (DeWitt and Graham 1973).

DeWitt, B. and N. Graham (Eds.) (1973). *The many-worlds interpretation of quantum mechanics.* Princeton: Princeton University Press.

Donald, M. (1995). A Mathematical Characterisation of the Physical Structure of Observers. *Foundations of Physics 25*, 529–571.

Donald, M. (1997). On Many-Minds Interpretations of Quantum Theory. Available at http://www.arxiv.org/abs/quant-ph/9703008.

Donald, M. (1999). Progress in a Many-Minds intepretation of quantum theory. Available at http://www.arxiv.org/abs/quant-ph/9904001.

Einstein, A. (1905). Zur electrodynamik bewegter körper. *Annalen der Physik*, 891–921. Translated in *The Collected Papers of Albert Einstein*, volume 2, J. Stachel (Ed.), Princeton, 1989.

Everett III, H. (1957). Relative State Formulation of Quantum Mechanics. *Review of Modern Physics 29*, 454–462. Reprinted in (DeWitt and Graham 1973).

Gell-Mann, M. and J. B. Hartle (1990). Quantum Mechanics in the light of Quantum Cosmology. In W. Zurek (Ed.), *Complexity, Entropy and the Physics of Information*, pp. 425–459. Redwood City, California: Addison-Wesley.

Ghirardi, G., A. Rimini, and T. Weber (1986). Unified dynamics for micro and macro systems. *Physical Review D 34*, 470.

Griffiths, R. (1984). Consistent Histories and the Interpretation of Quantum Mechanics. *Journal of Statistical Physics 36*, 219–272.

Haag, R. (1996). *Local Quantum Theory: Fields, Particles, Algebras.* Berlin: Springer-Verlag.

Halliwell, J. J. (1993). Aspects of the Decoherent Histories approach to quantum mechanics. In L. Diósi and B. Lukács (Eds.), *Stochastic Evolution of Quantum States in Open Systems and in Measurement Processes*, pp. 54–68. Singapore: World Scientific.

Heisenberg, W. (1971). *Physics and beyond: Encounters and Conversations.* New York: Harper & Row. Translated by A. J. Pomerans.

Holland, P. (1993). *The Quantum Theory of Motion.* Cambridge: Cambridge University Press.

Kaku, M. (1993). *Quantum Field Theory: a Modern Introduction.* Oxford: Oxford University Press.

Kent, A. (1998). Quantum Histories. *Physica Scripta T76*, 78–84. Available at http://www.arxiv.org/abs/gr-qc/9809026.

Ladyman, J. (1998). What is Structural Realism. *Studies in the History and Philosophy of Science 29*, 409–424.

Lewis, D. (1986). *On the Plurality of Worlds.* Oxford: Basil Blackwell.

Lockwood, M. (1989). *Mind, Brain and the Quantum: the compound 'I'.* Oxford: Blackwell Publishers.

Lockwood, M. (1996a). 'Many Minds' Interpretations of Quantum Mechanics. *British Journal for the Philosophy of Science 47*, 159–188.

Lockwood, M. (1996b). 'Many Minds' Interpretations of Quantum Mechanics: Replies to Replies. *British Journal for the Philosophy of Science 47*, 445–461.

Mermin, N. D. (1998). What is quantum mechanics trying to tell us? *American Journal of Physics 66*, 753–767.

Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973). *Gravitation*. New York: W.H. Freeman and Company.

Page, D. N. and W. K. Wootters (1983). Evolution without Evolution: Dynamics described by Stationary Observables. *Physical Review D27*, 2885–2892.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pearle, P. (1989). Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A 39*(5), 2277–2289.

Psillos, S. (1995). Is Structural Realism the Best of Both Worlds? *Dialectica 49*, 15–46.

Quine, W. v. O. (1966). Posits and reality. In *The Ways of Paradox and other essays*, pp. 246–254. Cambridge, Mass.: Harvard University Press.

Saunders, S. (1992). Locality, Complex Numbers and Relativistic Quantum Theory. *Philosophy of Science Association 1992 1*, 365–380.

Saunders, S. (1993). Decoherence, Relative States, and Evolutionary Adaptation. *Foundations of Physics 23*, 1553–1585.

Saunders, S. (1995). Time, Decoherence and Quantum Mechanics. *Synthese 102*, 235–266.

Saunders, S. (1996a). Comment on Lockwood. *British Journal for the Philosophy of Science 47*, 241–248.

Saunders, S. (1996b). Time, Quantum Mechanics and Tense. *Synthese 107*, 19–53.

Saunders, S. (1997). Naturalizing Metaphysics. *The Monist 80*(1), 44–69.

Saunders, S. (1998). Time, Quantum Mechanics, and Probability. *Synthese 114*, 373–404.

Stein, H. (1991). On relativity theory and openness of the future. *Philosophy of Science 58*, 147–167.

Tappenden, P. (2000a). Everett, Deutsch, doppelgangers and superposition. Forthcoming.

Tappenden, P. (2000b). Identity and Probability in Everett's multiverse. *British Journal for the Philosophy of Science*, 99–114.

Vaidman, L. (1998). On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory. *International Studies in Philosophy of Science 12*, 245–261. Available at http://www/arxiv.org/abs/quant-ph/9609006.

Wald, R. M. (1984). *General Relativity.* Chicago: University of Chicago Press.

Worrall, J. (1989). Structural Realism: the Best of Both Worlds. *Dialectica 43*, 99–124. Reprinted in *The Philosophy of Science*, D. Papineau (Ed.), Oxford University Press, 1996, pp. 139–165.

Zurek, W. H. (1991). Decoherence and the Transition from Quantum to Classical. *Physics Today*, 36–44.