

A Study of Time in Modern Physics

P. W. Evans

Doctor of Philosophy

The University of Sydney

2011

Declaration

This thesis is an account of research undertaken between February 2007 and August 2011 at the Centre for Time, Department of Philosophy, School of Philosophical and Historical Inquiry, University of Sydney, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Peter W. Evans

August, 2011

Acknowledgements

First and foremost I would like to extend much thanks to my supervisor, Prof. Huw Price, for both providing me the opportunity to complete this doctoral thesis on such a fascinating topic and for his sage advice throughout the process. I gratefully acknowledge the University of Sydney for the University Postgraduate Award that made this candidature possible and the Centre for Time, as well as the School of Philosophical and Historical Inquiry and the Department of Philosophy at the University of Sydney, for providing the support to attend and present at a range of conferences and workshops.

I thank both Dr. Guido Bacciagaluppi and Dr. Owen Maroney for providing co-supervision at various stages throughout my candidature. I would especially like to thank Dr. Kristie Miller for all the help she has provided and Assoc. Prof. Ken Wharton for some really stimulating discussions.

I extend a very special thanks to the fellow postgraduate students who provided endless support over the last few years discussing work, reading work and unwinding from work. Thank you very much Sam Baron, John Cusbert, Mikey Slezak and especially Karim Thébault; without your constant help this would have been a much more difficult experience.

I am grateful to all involved in the University of Sydney foundations of physics community, philosophy community and the various visitors that have passed through for many interesting and stimulating discussions.

I would also like to thank Wayne Myrvold, Dean Rickles and an anonymous examiner for comments and advice that have facilitated a vast improvement in this project.

Lastly, thanks to my family and friends, who have provided such great moral support.

Contents

Declaration	iii
Acknowledgements	v
Introduction	1
I Physical Theory and the Metaphysics of Time	7
1 The Picture of Time in Classical Mechanics	9
1.1 Introduction	10
1.2 Time in Newtonian mechanics	10
1.2.1 The spacetime formulation of Newtonian mechanics	14
1.3 Time in Lagrangian mechanics	16
1.4 Time in Hamiltonian mechanics	21
1.5 The Newtonian picture of time	26
2 Relativistic Constraints for a Naturalistic Metaphysics of Time	31
2.1 Introduction	31
2.1.1 Background	33
2.1.2 Outline of the chapter	35
2.2 Minkowski spacetime	35
2.3 Spacetime and reality	40
2.4 Objective temporal passage	44
2.5 Characterising time	48
2.6 The traditional debate constrained	52

3	Timelessness in Machian Gravity	59
3.1	Introduction	59
3.2	The Jacobian formulation of classical mechanics	60
3.3	Machian dynamics	63
3.4	Canonical quantum gravity	65
3.5	From timeless physical theory to timelessness	70
II	Quantum Theory and the Newtonian Picture of Time	77
4	Quantum Mechanics, EPR and Escaping Bell's Theorem	79
4.1	Introduction	80
4.2	Classical physics	81
4.3	The theory of quantum mechanics	84
4.4	The fifth Solvay conference, 1927	90
4.5	The EPR argument	93
4.6	Bell's theorem	97
4.7	EPRB and retrocausality	103
4.8	The lesser-of-two-evils	106
5	Retrocausality at No Extra Cost	109
5.1	Introduction	109
5.2	Foundations	111
5.2.1	The block universe model of time	111
5.2.2	The interventionist account of causation	111
5.3	Dismantling intuitions	114
5.3.1	Macroscopic intuitions, microscopic symmetry	114
5.3.2	The bilking argument	116
5.4	Keeping up appearances	121
5.5	A retrocausal picture	126

6 Causal Symmetry and the Transactional Interpretation	127
6.1 Introduction	127
6.2 The Wheeler-Feynman absorber theory of radiation	128
6.3 The quantum handshake	130
6.4 The transactional interpretation	132
6.5 Maudlin’s objection	136
6.6 Cramer defended	138
6.7 Maudlin’s experiment in four dimensions	141
6.8 Causal symmetry	145
6.8.1 Causality and determination	147
6.8.2 A tractable alternative	149
6.8.3 Classical tractability	150
6.9 Cramer’s missing structure and Maudlin’s misdirected metaphysics .	152
Summary	155
Bibliography	158

Introduction

This thesis is a study of the notion of time in modern physics, consisting of two parts. Part I takes seriously the doctrine that modern physics should be treated as the primary guide to the nature of time. To this end, it offers an analysis of the various conceptions of time that emerge in the context of various physical theories and, furthermore, an analysis of the relation between these conceptions of time and the more orthodox philosophical views on the nature of time. In Part II I explore the interpretation of nonrelativistic quantum mechanics in light of the suggestion that an overly Newtonian conception of time might be contributing to some of the difficulties that we face in interpreting the quantum mechanical formalism. In particular, I argue in favour of introducing backwards-in-time causal influences as part of an alternative conception of time that is consistent with the picture of reality that arises in the context of the quantum formalism. Moreover, I demonstrate that this conception of time can already be found in a particular formulation of classical mechanics. One might see that one of the central themes of Part II originates from a failure to heed properly the doctrine of Part I: study into the nature of time should be *guided* by modern physics and thus we should be careful not to *insert* a preconceived Newtonian conception of time unwittingly into our interpretation of the quantum mechanical formalism. Thus, whereas Part I is intended as a demonstration of methodology with respect to the study of time, Part II in a sense explores a confusion that can be seen as arising in the absence of this methodology.

To help clarify the general philosophical outlook that I will adopt here, let me begin by borrowing a distinction drawn by Wilfrid Sellars (1963, p. 19) between what he calls the *manifest image* and the *scientific image*. Sellars describes the manifest image as “the correlational and categorical refinement” of our crude perception of the world into a “framework of sophisticated common sense”; we can think of the manifest image as our refined intuitive picture of the world. In contrast, the scientific image of the world is “the image derived from the fruits of postulational theory construction”. Sellars views the manifest and scientific images as “two whole ways of seeing the sum of things” and his intention is “to bring them together in

a ‘stereoscopic’ view” of the world. We can employ this distinction for our present purposes and characterise this thesis as a study of the *picture of time* found within the scientific image of the world, as well as an examination of how the manifest image can be unwittingly subsumed into this scientific image.¹

With respect to this viewpoint, this thesis can be seen as embracing the *naturalistic metaphysics* of Ladyman and Ross (2007, Ch. 1): metaphysical inquiry is legitimate only when it “can be regarded as... [an] attempt to model the structure of objective reality” and “is motivated exclusively by attempts to unify hypotheses and theories that are taken seriously by contemporary science”.² While I adopt the position that contemporary science should be treated as an authority on the nature of reality, I stop short of advocating the structural realist program that Ladyman and Ross have in mind; indeed, I wish to abstain from any commitments concerning realism or anti-realism. In particular, I do not imagine time to be some objective feature of reality, independent of our descriptions of the world, such that we understand various scientific theories to be providing a different representation of this particular aspect of objective reality. Rather, by ‘reality’ here I simply mean our scientific image of the world, and thus this thesis is simply a study of how time is portrayed by some of our modern physical theories. By taking this stance, though, we do require a precise account of what we actually mean by ‘time’ in the context of different scientific theories.

To this end, I propose (in Chapter 2) a precise framework for characterising time that builds upon an analysis due to Rovelli (1995, 2004). According to Rovelli, the formal structure that we identify as time within a particular physical theory can be characterised in terms of its *attributes*; indeed, there are up to nine distinct attributes identified by Rovelli that we assign to the temporal structure of our various contemporary physical theories and folk concepts, including directionality, uniqueness and globality, amongst others. Rovelli proposes that our contemporary physical theories can be arranged in a hierarchical structure in which an increase in the universality of the theory corresponds to a decrease in the possible attributes that we can assign to the temporal structure of each theory. Thus various conceptions of time arising from various physical theories will consist of varying collections of

¹I leave to one side independent discussion concerning time in the manifest image.

²In contrast, metaphysical inquiry that proceeds without proper regard for science is labelled (somewhat pejoratively) *neo-scholastic metaphysics*.

temporal attributes; it is this notion of time that I will employ here.

At the core of this project is the authority of science as a guide to the nature of reality. As beings that have evolved as part of this reality, we have many firm intuitions that equip us to navigate successfully about the world. Modern science has taught us, however, that the world about which we navigate occupies a very small proportion of the universe: not only are we unable to venture to the spatial and temporal extremities of the universe, but our everyday lives are bereft of direct experience of its very small and very large scale structure. Thus while we might think that our natural place as inhabitants of this reality renders our intuitions finely attuned to the essential composition of the universe, we have today reached a scope of inquiry in modern physics where we are struck with the realisation that this is patently not the case. In spite of this realisation, many of our intuitions about the nature of reality are so engrained that we find it unthinkable that they might be called into question, or we simply fail to notice the presence of these intuitions dominating our scientific inquiries. This project explores the possibility that our intuitions concerning the nature of time are just such intuitions.

Let me turn now to a more detailed overview of the thesis. Part I, consisting of Chapters 1, 2 and 3, works from the premise that our best insight to the nature of time is attained through a study of the picture of time that emerges from physics. The physical theories on which I focus in Part I are theories of mechanics. My goal in Part I is to explore those aspects of the interpretation and formal mathematical machinery underpinning each physical theory that we usually identify as time and, through a consideration of the nature of this *temporal structure*, to examine and compare the corresponding picture of time within each theory. It is these pictures of time that I claim should be treated as our primary guide concerning the nature of time.

I begin in Chapter 1 with the familiar territory of analytical mechanics, including the Newtonian, Lagrangian and Hamiltonian formulations; commencing here serves as a clear introduction to the sort of analysis that I employ throughout Part I. The chapter begins by setting out the Newtonian picture of reality built upon the metaphysical notion of time as an external parameter ‘generating’ dynamical evolution; this is the Newtonian picture of time. Significantly, when we consider the geometric structure of both the Lagrangian and Hamiltonian formulations of mechanics we find a novel and interesting picture of reality arising. Despite this, the Newtonian

picture of time is usually thought to be the appropriate conception of time when considering the nature of dynamical evolution within the context of these formulations. We return to classical mechanics in Chapter 3 to add more to this story and also return to the discussion of the Lagrangian and Hamiltonian pictures in a retrocausal context in Chapter 6.

In Chapter 2 our study of the nature of time in physics is extended in two ways. Firstly, we move here beyond classical mechanics to consider the temporal structure of relativistic mechanics. Secondly, it is here that I connect this study with the philosophical literature concerning the nature of time. I introduce what I call the *traditional metaphysical debate* between opposing viewpoints on the essential features of time. At the core of this debate is whether time can be thought to *flow objectively*; that is, whether we imagine time to be *static* or *dynamic*. While much ink has been spilled debating this issue, I wish to address the problem employing the doctrine that physics should be the primary guide to the nature of time. To this end, I set out to examine the way in which considerations of the notion of time in both special and general relativity impinge on the metaphysical debate. In particular, I outline the constraints imposed by the temporal structure of relativity theory that the competing views of time comprising the traditional metaphysical debate must heed to remain within the bounds of a naturalistic metaphysics.

In Chapter 3 we turn our attention to the temporal structure of both a novel formulation of relational mechanics and the interpretation of quantum gravity that it motivates.³ Julian Barbour (1994a,b, 1999) develops a Machian formulation of general relativity that promotes a particular interpretation of canonical quantum gravity and then makes the claim that both theories are *timeless*. I introduce both of Barbour's theories, building upon the account of classical mechanics of Chapter 1, and I challenge his claim of timelessness: first, by identifying two different senses of timelessness that Barbour is using between his two theories; and, second, showing that we have reason to be suspicious of the claim that his Machian formulation of general relativity is in fact timeless. I utilise in this analysis my framework for characterising time from Chapter 2 to define the essential features of time. This concludes Part I.

³This chapter has developed from my contribution to the collaborative research paper Baron, Evans and Miller (2010). I thank my coauthors for the opportunity to reproduce some of this work in my thesis.

Part II, consisting of Chapters 4, 5 and 6, is an exploration of the interpretational difficulties that can arise when we fail to take seriously the doctrine of Part I concerning the nature of time. If our best insight into the nature of time is attained from the picture of time that emerges from a physical theory, then we must show caution when attempting to interpret the nonrelativistic quantum mechanical formalism in terms of the conception of time that fits most naturally in Newtonian mechanics; that is, the *Newtonian picture of time*. My aim in Part II then is to explore a ramification of eschewing the Newtonian picture of time through an examination of *retrocausality* as a solution to the interpretational difficulties of quantum mechanics. I argue that the main objections against including retrocausality in our quantum picture of reality are themselves built upon the Newtonian picture of time, and thus do not find immediate support from quantum mechanics itself.

We commence Part II in Chapter 4 where I introduce the formalism of nonrelativistic quantum mechanics alongside a somewhat historical dialectic of the development of its interpretation. I pay particular attention to the EPR argument and Bell's theorem and focus on the issue of nonlocality that arises in this context. I claim here that nonlocality creates difficulties for the interpretation of the quantum formalism due to an insistence on maintaining an overly Newtonian conception of time. I demonstrate that the introduction of retrocausal influences into the quantum picture of reality provides an *action-by-contact* explanation of this nonlocality alleviating this particular interpretational difficulty.⁴

In Chapter 5 I leave to one side the general project of examining the picture of time that arises from the formalism of physical theory and construct independent support for retrocausality through a more philosophical analysis. Much of the distaste surrounding the introduction of retrocausality in quantum mechanics stems from the intuition that causation must proceed from past to future, and that it is impossible to change the past. I address this challenge in Chapter 5 and demonstrate why we should not expect the former intuition to be relevant on an atomic scale and why we can maintain the latter intuition simply because retrocausality is not about 'changing' the past. I thus show that retrocausality cannot be ruled out on analytic grounds. The key to this argument is recognising that we perceive the world from a

⁴Some might wish to argue that an action-by-contact explanation by definition provides a *local* explanation. This depends upon what one means by locality. Discussion of this point can be found in Chapter 4, in particular §4.7.

particularly special vantage point: we are *agents* embedded in spacetime with a particular temporal perspective and a particular anthropological history. The picture of reality that arises from these considerations provides support for retrocausality as a solution to the interpretational difficulties of quantum mechanics.

In Chapter 6, the final chapter, we consider one of the most significant obstacles for retrocausal approaches to quantum mechanics in the form of the objection levelled at John Cramer's (1986) transactional interpretation of quantum mechanics by Tim Maudlin (2002). The transactional interpretation is a retrocausal model of quantum mechanics and Maudlin has developed an inventive thought experiment that he takes to pose a problem not only for the transactional interpretation, but also for retrocausality in general. I embark on an examination of the transactional interpretation to demonstrate that Maudlin's objection is indeed a problem for Cramer's theory but not a problem for retrocausality. The reason that Maudlin's objection fails to invalidate retrocausal theories in general is ultimately because his argument is grounded in the Newtonian picture of time. Recognising this fact renders the weaknesses of the transactional interpretation more clearly and exposes explicitly the way in which an overly Newtonian picture of time might be contributing to the difficulties we face in constructing a coherent interpretation of quantum mechanics.

The thesis concludes with a Summary in which I give an overview of the main results.

Part I

Physical Theory and the Metaphysics of Time

The Picture of Time in Classical Mechanics

We begin with an analysis of the picture of time that arises in the context of different physical theories. As mentioned in the Introduction, this first part of the thesis takes seriously the doctrine that modern physics should be treated as the primary guide to the nature of time. The goal here then is to explore the nature of time according to physics. There are three major parts to this analysis. The first part, which is the concern of this chapter, is an exploration of time in classical mechanics, which I characterise through the formal structure of Newtonian mechanics, Lagrangian mechanics and Hamiltonian mechanics. The aim here is to become better acquainted with analyses of the formal mathematical structure comprising a physical theory beginning with some of the more familiar classical physical theories. The second part, in the next chapter, is an analysis of the picture of time that arises from the geometrical structure of relativity theory, both the special and general theories. I also introduce at this point the traditional philosophical issues concerning the metaphysics of time. The goal of this analysis is to show that the picture of time that arises in relativity theory provides constraints on the metaphysical possibilities for time. The third and final part of this examination, in Chapter 3, addresses the picture of time that arises from both Julian Barbour's Machian formulation of general relativity and his interpretation of quantum gravity. These cases are particularly interesting due to Barbour's claim that these theories are *timeless*. I examine here the extent to which Barbour's interpretation of general relativity and quantum gravity can be seen as justifying the conclusion that we should think of the scientific image of the world as timeless.

1.1 Introduction

The aim of this chapter is to explore what I call the *Newtonian picture of time*. In short, we can think of the Newtonian picture of time as the notion that time is an independent and external parameter that generates dynamical evolution in physical systems. We begin in §1.2 with an exploration of the picture of time in Newtonian mechanics. I present an argument that the Newtonian conception of time can be seen as intimately linked with the mathematical formalism that underpins Newton's theory of mechanics: the calculus. We then move beyond Newton's formulation of mechanics to consider the tradition of *analytical mechanics*: a refined mathematical and geometrical formulation of classical mechanics. We first consider Lagrangian mechanics in §1.3 and then Hamiltonian mechanics in §1.4. In both cases I introduce the formal geometric structure of the theory and explore the picture of reality that arises in the context of each. What we find in each case is that the geometric structure of each theory provides a novel and interesting picture of reality in contrast to the Newtonian picture. Despite this, the Newtonian picture of time remains steadfastly attached to the picture of reality that arises from classical mechanics; this issue is discussed in §1.5. Let us begin with an introduction to Newtonian mechanics.

1.2 Time in Newtonian mechanics

In 1687, Isaac Newton (1962) published the first edition of his *Philosophiæ Naturalis Principia Mathematica*, in which he sets out, amongst other things, his theory of gravitation. The theory that is presented in the *Principia* is built upon Newton's famous three laws of motion as well as an adherence to mathematical principles in describing the motion of bodies through space and time. The resulting dynamical picture we call Newtonian mechanics and the goal of this section is to explore this dynamical picture of reality.

The modern geometrical (spacetime) formulation of Newtonian mechanics has been developed in more recent times to emphasise the similarities and differences between it and relativistic mechanics. Thus the mathematical formalism relevant to the temporal structure of Newtonian mechanics in its modern geometrical formulation is not open to interpretation in the same way as the formalism of the other

physical theories considered in Part I since it is motivated quite explicitly by the metaphysical notion of time at its core. It would be rather injudicious then to claim this formulation of Newtonian mechanics as an authority when considering Newtonian temporal structure. The mathematical formalism that we do find to underlie the metaphysical notion of time at the core of the Newtonian picture of reality is simply the calculus. I present the modern geometrical formulation of Newtonian mechanics below solely for the purposes of completeness and for comparison with the formalism introduced in the remainder of this part of the thesis. Before we consider this formalism, however, let us begin by considering the picture of time at the heart of Newtonian mechanics.

In the first few pages of the *Principia* Newton sets out his metaphysical view of *time, space, place* and *motion*, distinguishing between each concept as understood as “absolute, true, and mathematical” on the one hand, and “relative, apparent, and common” on the other. Concerning time, Newton famously says the following:

Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year. (1962, p. 6)

This picture of absolute time has come to be one of the defining features of Newton’s picture of reality. Although absolute time is often labelled an extraneous metaphysical assumption over and above the fundamentals of Newtonian mechanics (most notably by Mach (1960)), Arthur (1995) suggests that it is possible to understand absolute time as an integral element of Newton’s interpretation of the mathematical formalism which underpins the formulation of his theory: the calculus.

Central to the Newtonian picture is Newton’s theory of gravitation. The theory posits that the gravitational force that one body exerts on another is inversely proportional to the square of the distance between them. Thus if the dynamical behaviour of a system is completely determined by the gravitational forces between a collection of interacting bodies, then the key to determining this behaviour is the idea that a physical system can be described as a series of instantaneous spatial configurations embodying these relative distances. This then is the core of the Newtonian picture: a physical system is comprised of a collection of interacting bodies

(which can be treated as point particles of corresponding mass) that each have a definite spatial position at every instant of time. The dynamical behaviour of a Newtonian physical system is determined by the net force that acts on each body; forces are responsible for changes in momentum. To describe such dynamical behaviour requires the attribution of kinematical properties, such as velocity (momentum), to these interacting bodies, and to achieve this one needs to add to this picture an account of what it means for such a property to have an instantaneous value. This story, of course, is what is provided by Newton's calculus.

A *geometrical* relationship between a mathematical curve, its tangents and the area it bounds was well known in the 17th century (Boyer, 1970). It was also known that the velocity of a body in constant motion can be established mathematically via a ratio of a body's displacement to the time taken for the body to undergo this displacement, which is just the gradient of the straight line representing the motion of that body over time. Establishing a relationship between an *instantaneous* value of velocity on a *curve*, however, is problematic with respect to the indivisible quantities underlying the geometrical method. By extending this methodology and imagining such mathematical curves to be *continua* comprised of *infinitesimals*, the gradient at a point on the curve, and hence the instantaneous value of a kinematical variable fitting this mathematical description, can be determined theoretically. This extension to the geometrical method, known as the calculus, is a central feature of Newton's description of reality according to mathematical principles.

The conceptual leap involved in developing the calculus, and one which is of most importance for the present discussion, leads us from the geometrical interpretation of the mathematical formalism of the calculus to what we will call, following Arthur (1995), the *kinematical* interpretation: a mathematical curve is the trace of a moving point. Adopting this interpretation of the calculus can be seen as an attempt to find a physical grounding to the mathematically abstract notion of the infinitesimal, which leads us from the instantaneous to the continuous. If the motion of a body through space were described mathematically by a curve representing displacement as a function of time, then there exists a correspondence between, on the one hand, the mathematical trace of a moving point and, on the other, the motion of a body through space in a time that we conceive as the trace of a continually moving instant. Just as we can imagine a constant monotonic parameter as generating a parametric curve, the kinematical interpretation of the calculus suggests that

a constant monotonic time (in other words, an equably flowing time), generates the motion of bodies. This reading of the calculus emphasises that time must be unidirectional and acts as a parametric function on configurations of bodies.

Consider further that it must be the case that any empirical temporal measure of the dynamical behaviour of a system is relative to some given local motion that is putatively taken to be reliably periodic; that is, relative to a clock. Since there is no way to know whether any “relative time” such as this actually “flows equably”, we are led to the conclusion that some “absolute”, equably flowing time *must* be underlying physical dynamical systems (if we are to take this kinematical interpretation of the calculus seriously). Arthur’s kinematical interpretation suggests a quite plausible justification for Newton’s stubborn insistence on the metaphysical notion of absolute time in the *Principia* and also in his later works and correspondence.

The Newtonian picture of reality is built upon instantaneous configurations of bodies in an absolute spatial framework whose dynamical behaviour is determined by forces that act to change the momentum of these bodies with respect to an equably flowing and absolute time. This picture compels one to interpret the instantaneous state of a physical system as being in some sense ‘generated’ by its antecedent state, and also itself ‘generating’ the subsequent state. Let us call this Newtonian picture of determination the *generative* picture. The explicit adherence to mathematical principles in describing physical systems contributes to the view that reality evolves mechanistically in time along these lines. This is a crucial element of the Newtonian picture of time, and one which will be in our sights for the remainder of this thesis.

For now let us turn our attention to the modern geometrical formulation of Newtonian mechanics built upon the metaphysical notion of absolute time. As mentioned above, this formalism is presented here to parallel the remainder of the analysis of Part I, especially our concern with the four dimensional structure of relativity theory in Chapter 2.¹

¹The exposition here mostly follows Friedman (1983) and Rodrigues, de Souza and Bozhkov (1995). See also Schutz (1980) for a great introduction to the geometry employed throughout the remainder of this chapter and beyond.

1.2.1 The spacetime formulation of Newtonian mechanics

Newtonian spacetime can be represented by a differentiable, four dimensional manifold, \mathcal{M}^4 , where each point $p \in \mathcal{M}^4$ is interpreted as an *event* in spacetime. We define on \mathcal{M}^4 an affine connection, D , with the property that it determines a curvature tensor field that vanishes for all $p \in \mathcal{M}^4$; we say that Newtonian spacetime is *flat*. We define further a one-form field, $dt : T\mathcal{M}^4 \rightarrow \mathbb{R}$ that is the gradient of a continuous, differentiable *time function*², $t : \mathcal{M}^4 \rightarrow \mathbb{R}$; we represent absolute time by the one-form field dt . A vector $u_p \in T_p\mathcal{M}^4$ is said to be *spacelike* just in case $dt_p(u_p) = 0$ and *timelike* just in case $dt_p(u_p) \neq 0$, whereby it is *future directed* if $dt_p(u_p) > 0$ and *past directed* if $dt_p(u_p) < 0$. In this way, dt defines a ‘direction’ to absolute time in the tangent bundle $T\mathcal{M}^4$. The number $t(p)$ represents the *time* for each event $p \in \mathcal{M}^4$ and $|t(q) - t(p)|$ represents the *temporal interval* between any two events $p, q \in \mathcal{M}^4$; we say two events are *simultaneous* if $t(p) = t(q)$.

For each $p \in \mathcal{M}^4$, the set $S_p = \{q \in \mathcal{M}^4, t(q) = t(p)\}$ of all events simultaneous with p defines a three dimensional submanifold of \mathcal{M}^4 (the set of all tangent vectors to this set, $T_p S_p$, coincides with the set of all spacelike vectors at p). We can ensure that each submanifold is a Euclidean (flat) space, E^3 , by requiring compatibility between the one-form field, dt , and the affine connection, D : $D(dt) = 0$. A metric tensor can be defined on \mathcal{M}^4 which collapses into the usual Euclidean metric when restricted to the set of spacelike vectors at p for all $p \in \mathcal{M}^4$. Thus absolute time and the affine connection give a unique foliation of the Newtonian spacetime manifold into a continuous class of three dimensional Euclidean submanifolds which we interpret as the instantaneous spatial configurations at the core of the Newtonian picture.

Rather than \mathcal{M}^4 , we could have represented the structure presented thus far as a fiber bundle $E^3\mathbb{R}$ whose base space is absolute time and whose fibers are three dimensional spaces. This particular fiber bundle structure emphasises the fact that there is no natural relation between points of space at different times (points on different fibers) (Schutz, 1980); we must insert such a relation by hand. We do so by introducing a vector field V such that $dt(V) = 1$ and $D(V) = 0$; the curves γ defined by V (i.e. those γ along which $T_\gamma = V$) are timelike by the former condition

²In fact, dt represents the gradient of a whole set of time functions $\{t+b\}$ for arbitrary constants b .

and are geodesics by the latter. We let these curves determine the trajectories that define absolute rest, such that two events $p, q \in \mathcal{M}^4$ occur at the same place in space if and only if they lie on the same curve γ . This yields a notion of absolute space in Newtonian spacetime. We interpret any future directed geodesic curve as a possible spacetime trajectory for a free body in Newtonian mechanics, and any future directed timelike curve as a possible trajectory for *any* body in Newtonian mechanics. The trajectory of every Newtonian body is parametrised by absolute time.

Newton's Law of Inertia³ is a consequence of the connection, D : the trajectories of free bodies satisfy the geodesic equation of motion $D_{T_\gamma} T_\gamma = 0$, which, in an inertial coordinate system x_i , takes the simple form

$$\frac{d^2 x_i}{dt^2} = 0. \quad (1.1)$$

Newton's second law of motion states that the change in a body's momentum with respect to time is proportional to the net force acting upon that body; that is,

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}, \quad (1.2)$$

where the vector \mathbf{F} represents the force acting on the body and $\mathbf{p} = m\dot{\mathbf{x}}$ represents the momentum, with m representing the mass of the body and $\dot{\mathbf{x}}$ its instantaneous velocity. The dynamical behaviour of a physical system of N bodies is thus described by a system of $3N$ second-order differential equations and, given the instantaneous position and velocity of each of these bodies, Newton's second law comprises a well-posed initial value problem that provides a unique description of the subsequent behaviour of the system in question.

The absolute temporal structure of Newtonian spacetime together with Newton's Law of Inertia imply that a free body will travel equal distances in equal intervals of time and, thus, the "equable flow" of absolute time is ensured. It is also apparent from the definition of dt that absolute time is unidirectional so long as t is monotonic, but whether time flows towards the past or towards the future is a matter

³In Newton's words:

Every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed upon it. (1962, p. 13)

of definition; Newtonian mechanics is *time reversal invariant*. In addition, no three dimensional submanifold $E^3 \subset \mathcal{M}^4$ can be considered special in any sense; all Euclidean ‘slices’ of Newtonian spacetime are considered to be on an equal footing. These elements contribute to the Newtonian picture of time as an external parameter with a constant flow that determines evolution in a generative fashion. Let us now consider the nature of time in Lagrangian mechanics.

1.3 Time in Lagrangian mechanics

Newtonian mechanics motivates a picture of reality in which the dynamical behaviour of a body is governed by the sum total of forces to which it is subject. This conceptual framework for describing physical systems, however, is limited in practice since many physical systems are characterised by complicated ‘constraint’ forces that compromise the tractability of dynamical models⁴, especially since ‘force’ (as well as ‘momentum’) is a vectorial quantity (having both magnitude and three dimensional direction). A more general schema for modelling dynamical systems, which we call Lagrangian mechanics, was developed by Joseph-Louis Lagrange (1753–1828) in his 1788 work *Mécanique Analytique*. Lagrange achieves this generality in his system of mechanics by replacing the vectorial quantities underlying Newtonian mechanics with scalar quantities relating to the energy of the physical systems in question; this enables the mathematical formalism describing dynamical behaviour to be independent of the coordinate system required for vectorial descriptions.

As calculational tools for determining or predicting the behaviour of dynamical systems, Newtonian and Lagrangian mechanics yield equivalent empirical results (despite differing in the simplicity with which the two treatments handle complicated systems). However, the shift in focus from vectorial force in Newtonian mechanics to scalar energy in Lagrangian mechanics creates scope for divergent physical interpretations of the formalism of each theory, especially with respect to the picture of time. The goal of this section is to explore the picture of time that arises from the formalism of Lagrangian mechanics. Let us begin by motivating Lagrange’s original formulation of his mechanics.⁵

⁴Think of the forces constraining the motion of a bead sliding down a helical wire; we call these forces *kinematical constraints*.

⁵The formalism in this section and the next is taken mostly from Lanczos (1970) and Belot (2007).

Consider once again the initial value problem that comprises Newton's second law of motion (1.2). We noted above that this law reduces to a set of $3N$ second-order differential equations for a system of N bodies in three dimensional space. We can thus completely describe the motion of these bodies using $6N$ independent variables, and we take these in Newtonian mechanics to be the Cartesian position coordinates, x_i , and their first time derivatives, the velocities, \dot{x}_i . In Lagrangian mechanics we move to a description of dynamical behaviour in terms of a set of *generalised coordinates*, q_i , and their first time derivatives, the *generalized velocities*, \dot{q}_i ; we do so by stipulating general transformation equations that express the Cartesian coordinates as arbitrary functions of the generalised coordinates. The generalised coordinates are often chosen to embody any kinematical constraints on the system.

Consider also at this point d'Alembert's principle: the total virtual work, δW , done by an arbitrary virtual displacement of a body in three dimensional space, $\delta \mathbf{x}$, which is impelled to move by an impressed force, \mathbf{F} , augmented by the inertial force of that body's mass, $m\ddot{\mathbf{x}}$, is zero (if the virtual displacement is in harmony with any kinematical constraints). The essence of d'Alembert's principle is that the dynamical behaviour of a body can be modelled by a static system in which there is equilibrium between impressed and inertial forces; the work done by the inertial forces associated with the motion of a body balances the work done by the corresponding impressed forces. A sum over bodies gives d'Alembert's principle for a many-body system:

$$\delta W = \sum_i (\mathbf{F}_i - m_i \ddot{\mathbf{x}}_i) \cdot \delta \mathbf{x}_i = 0. \quad (1.3)$$

D'Alembert's principle is equivalent to the Newtonian equations of motion. If we now apply the above transformation equations between the generalised and Cartesian coordinates, we can rewrite the virtual displacement $\delta \mathbf{x}$, impressed force \mathbf{F} and inertial force $m\ddot{\mathbf{x}}$ in terms of the generalised coordinates. This yields the Newtonian equations of motion in an arbitrary coordinate system.

Moreover, we can move from the vectorial quantities \mathbf{F} and $\ddot{\mathbf{x}}$ to the scalar quantities, V , potential energy and, T , kinetic energy so long as we can represent force as the gradient of a scalar potential energy function, $\mathbf{F} = -\nabla V$, and by employing the relation between the sum of inertial forces and the kinetic energy of the system. When both of these relations are expressed as functions of the generalised coordi-

nates and substituted into (1.3), we find (after a little rearranging) that a system of second-order differential equations arises from d'Alembert's principle as a consistency constraint on the possible dynamical behaviour of a mechanical system.⁶ These constraint equations, known as the *Euler-Lagrange equations*, have as their argument a scalar function $\mathcal{L} = T - V$ known as the *Lagrangian*:

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0. \quad (1.4)$$

As it happens, this derivation of the Euler-Lagrange equations conceals the fact that the two balancing forces of d'Alembert's principle, the impressed forces and the inertial forces, are very different in their nature (Lanczos, 1970, p. 111). The work done by the impressed forces is a function of the potential energy only, which is a characteristic of a system as a whole, while the work done by the inertial forces is a sum over the kinetic energy of each individual body of the system. In this way, we interpret the Euler-Lagrange equations thus derived as constraint equations that a system must obey *at each moment in time*, as opposed to a *global constraint* on the dynamical behaviour of the system. There is, however, a more powerful method for deriving (1.4) that licenses an interpretation of the Euler-Lagrange equations as just such a global constraint.

We can consider the work done by both the impressed and inertial forces on an equal footing if we integrate the expression for the work done, (1.3), over a definite interval of time. This move converts d'Alembert's principle into a *variational principle* of mechanics; this principle is known as Hamilton's principle. It can be shown that when we integrate δW between two temporal limits, t_a and t_b , provided we completely specify the state of the system at the temporal boundaries, the expression for the work done simplifies into a variation of a definite integral of the Lagrangian function from above:

$$\int_{t_a}^{t_b} \delta W dt = \delta \int_{t_a}^{t_b} \mathcal{L} dt = 0. \quad (1.5)$$

It is a general mathematical result in the calculus of variations that the necessary and sufficient condition for an integral such as this to be stationary is that the Euler-Lagrange equations, (1.4), be satisfied. Thus we find that we are able to interpret the

⁶This derivation assumes that there are no dissipative forces at play in the system.

Euler-Lagrange equations derived from Hamilton's principle as a global constraint on the dynamical behaviour of a system. To assist in understanding exactly what this means for the physical picture of Lagrangian mechanics, let us consider Hamilton's principle geometrically.

Consider the $3N$ dimensional manifold, Q , that is coordinatised by the generalised coordinates q_i of a system of N bodies. Each point in Q represents a possible set of values for the q_i , which together specify a configuration of the system; thus we call Q *configuration space*. Since the Lagrangian is a function of both the potential energy, which takes configurations as inputs to give a scalar, $V : Q \rightarrow \mathbb{R}$, and the kinetic energy, which takes both the generalised position and velocities as inputs to give a scalar, $T : TQ \rightarrow \mathbb{R}$, the Lagrangian is thus a scalar function on the tangent bundle TQ , $\mathcal{L} : TQ \rightarrow \mathbb{R}$. The dynamical behaviour of a physical system can be described as a curve γ through configuration space; for some closed interval $[a, b] \in \mathbb{R}$, $\gamma : [a, b] \rightarrow Q$ with endpoints $\gamma(a)$ and $\gamma(b)$. Each curve is parametrised by time, t , with $\gamma(a)$ the configuration of the system at time t_a and $\gamma(b)$ the configuration of the system at time t_b . At each point along the curve $\gamma(p)$, $p \in [a, b]$, there is a definite value for the Lagrangian $\mathcal{L}(q_i(t), \dot{q}_i(t))$ ($q_i, \dot{q}_i \in T_{\gamma(p)}Q$) and thus the Lagrangian function is time integrable along the curve from the endpoints t_a and t_b . This integral, i.e. the definite integral of \mathcal{L} in (1.5), we call the *action*.

We can now state Hamilton's principle in the following way: physically realisable dynamical behaviour is described by curves in configuration space for which the action becomes stationary for arbitrary possible variations of the configuration of the system, provided the initial and final configurations of the system are prescribed. Thus while there might be many curves between two points in Q , only one path (given a suitable topology) will be a critical (or extremal) point of the action. The curves picked out by Hamilton's principle are just those curves that satisfy the Euler-Lagrange equations (Belot, 2007, p. 145). Significantly, if the initial and final configurations of the physical system are not completely specified, we cannot derive (1.5) and thus Hamilton's principle fails.

The picture of reality that follows from the formalism of Lagrangian mechanics is thus rather different from the Newtonian picture we considered above. Rather than a set of forces that act on the individual bodies of an instantaneous configuration that affect a change in momentum with respect to time for each body, and thus generate the dynamics of the system, we have a global principle of critical action that acts on

the set of curves between two points in the configuration space of the system which determines dynamical behaviour. Each individual body of the system no longer has the sort of significance it had in Newtonian mechanics; the physical system is treated as a whole in Lagrangian mechanics. Moreover, the Lagrangian formalism expands the extension of “the physical system”: the objects to which Hamilton’s principle refers are curves in configuration space and these curves, parametrised by time, are four dimensional ‘stacks’ of the three dimensional configurations (or, equivalently, the configurations are three dimensional ‘slices’ of four dimensional curves). Since physically realisable dynamics is defined by a stationary action, and the action is a property of a four dimensional curve, the dynamical behaviour of a physical system according to Lagrangian mechanics is determined across a four dimensional structure. More specifically, both the initial configuration of the system and the final configuration of the system are equally influential in determining dynamical behaviour. Given the initial and final configurations, the single scalar function \mathcal{L} determines the entire dynamics of the system.

The differences between the Lagrangian and Newtonian pictures is obscured by the fact that each formalism yields equivalent results concerning the description of physical systems (at least in those cases where both are applicable). To emphasise these differences once more, recall the interpretation of Newtonian mechanics above in which an instantaneous configuration of a physical system (with specified velocities) can be thought as in some sense generated by its antecedent state and also as generating the subsequent state. According to such a view nature can be imagined as a linear computational machine which takes some instantaneous state of a system as input and gives the state of the system at the following instant as output. In contrast, the Lagrangian formalism requires two temporal boundary conditions to produce a solution; thus the initial data of Lagrangian mechanics would render such a linear machine underdetermined, with an initial configuration providing only half the data required. Nature according to the Lagrangian picture of reality is more like a four dimensional analogue to the three dimensional determination of an electric field between two charged plates: once the charge distribution of the two plates is specified, the electric field is determined by these distributions and the laws of

electrostatics.⁷

Despite the differing pictures of the dynamical structure underpinning our reality between Lagrangian and Newtonian mechanics, there does not seem to be a very significant departure in the ordinary interpretation of time in the Lagrangian schema from the Newtonian picture of time. Time in Lagrangian mechanics (the parametrisation of the physically realisable trajectories) is imposed as an external, constant monotonic parameter that generates dynamical evolution. Thus Newton’s appraisal that “time, of itself, and from its own nature, flows equably” remains consistent with picture of time in the Lagrangian schema. The most significant difference, however, when compared to the Newtonian picture of absolute time is that any such ‘flow’ of Lagrangian time is more naturally thought to be constrained in a rather interesting and novel way: at two temporal boundaries. We discuss this further in §1.5; for now, though, let us consider time in Hamiltonian mechanics.

1.4 Time in Hamiltonian mechanics

Although the Lagrangian formulation of Newtonian mechanics greatly simplifies the mathematical description of physical systems subject to kinematical constraints, the equations of motion which describe dynamical behaviour, the Euler-Lagrange equations, remain second-order differential equations. The Irish mathematician William Rowan Hamilton (1834), in his publication *On a General Method in Dynamics*, constructs an elegant reformulation of the Lagrangian equations of motion as a set of first-order differential equations. This new schema for describing the dynamical behaviour of physical systems we call Hamiltonian mechanics. The goal of this section is to explore the picture of time that arises from the Hamiltonian formalism.

Consider again the Euler-Lagrange equations (1.4). We are able to eliminate the second order derivative in the first term of this expression simply by defining a set of new variables,

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}. \quad (1.6)$$

We call p_i the *generalised momenta* (since in Cartesian coordinates $\frac{\partial \mathcal{L}}{\partial \dot{q}_i} = m\dot{x}_i$).

⁷A clutch of interesting issues arise at this point concerning the nature of causality and the four dimensional structure of reality when one looks a little closer at terms such as “generate” and “determine”. We will meet these issues again in more depth in Part II, particularly Chapter 5.

With the introduction of this new set of variables, the Euler-Lagrange equations take on a particularly simple form,

$$\dot{p}_i = \frac{\partial \mathcal{L}}{\partial q_i}. \quad (1.7)$$

Thus where we once had n second-order equations (1.4) (corresponding to $3N$ generalised coordinates describing an N -body system), we immediately obtain $2n$ first-order equations (1.6) and (1.7). Moreover, by introducing the new set of variables in this way, we are now able to apply a *Legendre transformation* to the Lagrangian function using these new variables.⁸ Doing so defines a new function, H , which we call the *Hamiltonian*,

$$H = \sum_{i=1}^n p_i \dot{q}_i - \mathcal{L}. \quad (1.8)$$

Just as $\mathcal{L} = T - V$, (1.8) renders $H = T + V$. Since we are able to solve (1.6) for the \dot{q}_i , we can express the Hamiltonian purely as a function of the q_i and p_i . The Legendre transformation then allows us to reexpress (1.6) and (1.7) as a function of H :

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \quad (1.9)$$

These equations are known as Hamilton's equations, or the *canonical equations*. Hamilton's equations are completely equivalent to (1.6) and (1.7); all we have done here is define a new set of variables and rewrite the Euler-Lagrange equations of motion in a novel mathematical form. Nonetheless, since we can express H purely as a function of the q_i and p_i , Hamilton's equations isolate the derivatives with respect to time to one side of the equations, thus greatly simplifying the problem of determining dynamical behaviour.

We have seen here that we can derive Hamilton's equations through a Legendre transformation of the Euler-Lagrange equations, which themselves arise as a constraint of a variational principle (1.5). The duality of the Legendre transformation suggests, however, that Hamilton's equations need not be thought a mere reexpression of the more primitive Euler-Lagrange equations. In fact, Hamilton's equations can also be derived themselves as constraint equations of this variational principle. We can use (1.8) to construct the variational principle (1.5) as a function of the

⁸We also require that we can rewrite (1.6) to get $\dot{q}_i(q_i, p_i)$.

Hamiltonian,

$$\delta \int_{t_a}^{t_b} \left[\sum_{i=1}^n p_i \dot{q}_i - H(q_i(t), p_i(t)) \right] dt = 0; \quad (1.10)$$

the new action integral we call the *canonical integral*. It just so happens that the new variables we have introduced into the Hamiltonian schema p_i , while defined as functions of the q_i and \dot{q}_i , behave under variation as if they were a second set of independent variables. This means that we can require of our new variational principle over the Hamiltonian function that it assume a stationary value for arbitrary variations of *both* the q_i and p_i . The constraint equations that arise from this new variational problem are just Hamilton's equations, since we have twice the number of independent variables and thus twice the number of differential equations constraining the dynamics; but, of course, now these constraint equations are first-order differential equations. Moreover, the form of the canonical integral makes apparent that each q_k is associated with its own p_k , and for this reason the p_k can be referred to as the *conjugate momenta*.

The introduction of a new set of independent variables is the significant move that sets apart the Hamiltonian picture of reality; let us consider this picture through the geometric structure of the Hamiltonian formalism. To begin with, we can characterise the generalised momenta as a geometric object in terms of the Lagrangian configuration space Q . Since we define $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$, under a change of coordinates on Q , we find that the momenta transform as a covariant quantity. Thus for each position $q \in Q$ we can represent the momentum of a physical system as a one-form on Q that lives in the cotangent space, $p \in T_q^*Q$. The cotangent bundle T^*Q is then the set of all pairs (q, p) with $q \in Q$ and $p \in T_q^*Q$; we define $\Gamma := T^*Q$ and call this the *phase space*. This space is the arena in which Hamiltonian mechanics describes dynamics. Like the tangent bundle TQ , the cotangent bundle T^*Q has dimension $2n$ ($6N$), which in this case corresponds to n position coordinates, q_i , and n momentum coordinates, p_i .

The coordinates q_i and p_i are not unique. While it is the case that we are able to transform the generalised coordinates that describe a physical system arbitrarily, it is not the case that any arbitrary coordinates will yield an expression for the Hamiltonian that renders an action integral in the form of the canonical integral, and thus it is not the case that any arbitrary coordinates will generate Hamilton's equations of motion. It turns out that requiring of any coordinate transformation

that it preserve the form of the canonical integral (and thus Hamilton's equations) is equivalent to requiring the existence of a coordinate independent one-form, θ , on Γ . Transformations of this sort we call *canonical transformations*, and we call θ the *canonical one-form*. The canonical one-form, as a one-form on Γ , lives in the cotangent space $T^*(\Gamma)$ and is a real valued function on vectors in the tangent bundle of the phase space; $\theta : T(\Gamma) \rightarrow \mathbb{R}$. By taking the exterior derivative of the canonical one-form, we define the *canonical two-form*, or *symplectic form*, on Γ , $\omega = d\theta$, which endows phase space with a particular geometric structure known as a *symplectic* structure; (Γ, ω) defines a symplectic manifold. Thus we find that dynamical behaviour as described by Hamilton's equations is encoded in the geometry of the cotangent bundle of the configuration space Q .

The symplectic form, $\omega : T(\Gamma) \times T(\Gamma) \rightarrow \mathbb{R}$, can be cast in a role on phase space similar to the role that a metric plays on a Riemannian manifold: ω provides an invertible one-to-one mapping between vectors in $T(\Gamma)$ and one-forms in $T^*(\Gamma)$. This is significant for our current purposes for the following reason. Recall that the Hamiltonian, H , is a scalar function of the generalised coordinates and momenta; $H : (q, p) \in \Gamma \rightarrow \mathbb{R}$. Thus the gradient of this function, dH , is a one-form on Γ living in $T^*(\Gamma)$. The symplectic form ω and the one-form dH then implicitly define a unique vector field X_H on the phase space: $\omega(X_H, \cdot) = dH$.

All we have done here is provide a novel reformulation of the Lagrangian formalism and we find ourselves waist deep in geometric structure. Let us begin to unpack this geometry with one eye on the Hamiltonian picture of reality we wish to unearth. Just as we imagine a point in the configuration space Q as representing one possible three dimensional spatial configuration of a physical system, we likewise imagine a point in the phase space T^*Q as representing a possible three dimensional configuration with a possible set of momenta associated with each body of the system. Thus we can think of phase space as a space of possible initial data for a dynamical problem of Newtonian mechanics. The Hamiltonian function $H = T + V$ can be thought straightforwardly to provide the total energy of a physical system at one of these initial data points. The vector field X_H generates an \mathbb{R} -action, and thus a *flow* $\{\Phi_t\}$, on the phase space such that, given a point $(q_0, p_0) \in \Gamma$, there is a unique integral curve $\gamma_{(q_0, p_0)}$ passing through that point. Since the vector field X_H is determined by the gradient of the Hamiltonian dH (which encodes information about the energy of a system) and the symplectic geometry derived from ω (which

embodies Hamilton's equations of motion) the integral curves of X_H describe the physically realisable trajectories that a physical system can trace through the phase space; these trajectories are simply the solutions to Hamilton's equations. Each Φ_t implements time evolution of some set of initial data: Φ_t maps a particular state of a physical system to the state into which it dynamically evolves after time t . Thus the triple (Γ, ω, H) completely specifies the dynamical behaviour of a physical system.

The Hamiltonian formulation of mechanics is formally equivalent to both the Lagrangian and Newtonian formulations (in those cases where each is applicable). However, the picture of reality that arises from Hamiltonian mechanics differs ever so slightly from the Lagrangian picture. In both cases, time enters the picture as the parameter of integration in the variational principle. The significance of the variational principle in the Lagrangian picture is to isolate those paths through configuration space that represent physically realisable trajectories for physical systems. Time in the Lagrangian picture is embodied in the parametrisation of these paths, and this parametrisation is constrained by the two temporal endpoints. In comparison, the variational principle in the Hamiltonian picture is responsible for the symplectic geometry of phase space and it is this geometry that ensures the integral curves of the Hamiltonian vector field correspond to physically realisable trajectories. Time in the Hamiltonian picture is again embodied in the parametrisation of these curves, although in this case the parametrisation is constrained by the flow $\{\Phi_t\}$ generated by X_H . While the interpretation of the points $(q, p) \in \Gamma$ as initial data points for a dynamical problem of Newtonian mechanics suggests a picture of reality for Hamiltonian mechanics that is similar to the Newtonian picture of reality, where each instantaneous configuration generates the next in turn, the fact that temporal evolution is constrained by X_H , which is ultimately a function of the variational principle (1.10), indicates that the Hamiltonian picture of reality also borrows elements of the Lagrangian picture, where the dynamical behaviour of a physical system is represented within a four dimensional stack of configurations.⁹

⁹There is a formal equivalence between solutions to the variation problem in Lagrangian configuration space and the symplectic geometry (Γ, ω) (Belot, 2007, p. 146). We can construct a space, \mathcal{S} , in which each point corresponds to a particular path in the Lagrangian configuration space with stationary action. This correspondence can be employed to coordinatise \mathcal{S} , and thus give \mathcal{S} a manifold structure with the same dimension as TQ . Moreover, it can be shown that there exists a symplectic two-form on \mathcal{S} and that it follows from this that \mathcal{S} paired with this symplectic two-form is a symplectic geometry isomorphic to (Γ, ω) .

One might ask at this point, however, whether we could dilute this variational flavour of the Hamiltonian schema by taking Hamilton's equations as basic. Since the Legendre transformations are completely symmetric there is no requirement that we must take the Euler-Lagrange formulation of mechanics as primary; we have seen above that we can formulate Hamilton's equations directly without the Lagrangian equations nor the Legendre transformations. However, I emphasise that to do so we must produce a new action integral in terms of an extended set of independent variables and subject it once again to a variational principle (1.10); Hamilton's equations become the conditions for a stationary action integral under arbitrary variations which are once again constrained by initial and final boundary conditions. Regardless then of how we might construct the geometry of Hamiltonian phase space, we find that our bedrock consists of a variational principle.

Once again, despite yet another interesting variation of the picture of dynamical structure underpinning our reality, there is no real departure in the ordinary interpretation of time in the Hamiltonian schema from the Newtonian picture of time. Time in Hamiltonian mechanics (again, the parametrisation of the physically realisable trajectories) is imposed as an external, constant monotonic parameter that generates dynamical evolution. The flow of this parameter again remains equable, but in this case the Hamiltonian flow is constrained by the geometry of phase space which encodes Hamilton's equations. Let us address these differing pictures of time in more detail.

1.5 The Newtonian picture of time

The picture of reality that arises from Newtonian mechanics portrays time as an independent and external parameter that generates dynamical evolution in physical systems in step with its equable flow. The nature of Newtonian determination compels us towards a generative picture of reality. The combination of these views is what I have called the Newtonian picture of time. While this Newtonian picture is in some sense crucial to understanding the Lagrangian and Hamiltonian pictures of reality, in another sense there is something a little unnatural about imposing it on the Lagrangian and Hamiltonian schema. Let us consider first the Lagrangian schema.

In the Lagrangian schema dynamical behaviour is determined by a variational

principle containing an action integral constrained at two temporal endpoints. Thus given the initial and final configuration of a particular physical system, the solution to the variational problem provides a sequence of configurations through which a physical system must pass to evolve from the initial to the final state. The temporal structure of the Lagrangian schema is built upon the action integral: not only does this integral require stipulation of the time at which the system is in both the initial and final configurations, we must also impose a constant monotonic parameter of integration that we identify as time. Thus time in the Lagrangian schema is explicitly an external parameter with an equable flow. In this sense, the picture of time in the Lagrangian schema is like Newtonian time.¹⁰

However, the nature of the variational problem of Lagrangian mechanics indicates that the Lagrangian picture is in a sense at odds with the Newtonian picture of reality as a linear generative computational machine. While it is a brute fact of the Lagrangian configuration space that it is possible in principle to uniquely specify a physically realisable trajectory given data consisting of a single configuration with corresponding velocities (and thus data at a single time), it seems unnatural to imagine the Lagrangian formalism as suggesting a picture of reality as a ‘black box’ data processor embodying the dynamical laws of nature that takes the present time instant as input and delivers the next time instant in sequence. Instead the Lagrangian conception of determination has a *teleological* nature: the evolution of a physical system is determined by fixed boundary conditions at both temporal endpoints. Taking this Lagrangian picture seriously would jeopardise this element of the Newtonian picture of time. (In Part II we will connect this Lagrangian picture with the nature of retrocausality.)

Although we find scope to compose a novel conception of time from the Lagrangian picture, the Newtonian picture of time remains firmly attached to its usual interpretation. For the most part this is because we can recover the Newtonian picture from the Lagrangian formalism. Using Hamilton’s principle we can obtain the Euler-Lagrange equations which under appropriate conditions will yield the Newtonian laws of motion. *Given these equations* we are free to imagine the evolution of the system as generated by the laws of motion. The significance of the route we must take to derive these equations (from a variational principle containing an action in-

¹⁰In Chapter 3 we will meet a formulation of classical mechanics that is similar to the Lagrangian schema except it does not require the imposition of an external time with constant flow.

tegral constrained at two temporal endpoints) appears to be overlooked within the usual interpretation, perhaps because the Lagrangian formulation of mechanics has largely been viewed as merely a mathematical trick that simplifies the calculations in dynamical problems. Thus despite the potential for a novel and interesting conception of time, the standard reading of the Lagrangian formalism appears wedded to the Newtonian picture of time.

A similar story can be told with respect to the Hamiltonian picture of reality. The temporal structure of the Hamiltonian schema is built into the geometry of the phase space. The dynamical behaviour of physical systems is determined by the vector field X_H which, recall, is defined by the gradient of the Hamiltonian dH , encoding information about the energy, and the symplectic geometry derived from ω , embodying Hamilton's equations of motion. X_H generates a flow $\{\Phi_t\}$ and, given any particular configuration, the flow determines a sequence of configurations that the system must pass through during its evolution in time; these are simply the integral curves of X_H . The symplectic geometry of phase space ensures a *unique parametrisation* of these integral curves and thus the way in which a system evolves through time is uniquely defined by the geometry of phase space. Given some initial data for a system then, the Hamiltonian schema is able to determine a unique dynamical evolution for that system, and thus the Hamiltonian schema seems to a certain extent commensurate with the generative Newtonian picture of determination.

It is contestable, however, whether or not this is the most natural interpretation of the formalism. The Hamiltonian vector field X_H is determined, in part, by the symplectic geometry of phase space, and this geometry encodes Hamilton's equations of motion. These equations of motion, like the Euler-Lagrange equations, are established via a variational principle containing an action integral between two temporal endpoints. Thus while the geometry of phase space (in particular, the vector field X_H) suggests that the generative picture of reality might naturally arise from the Hamiltonian schema, the origin of this geometry in a variational principle suggests a picture more like the Lagrangian picture. We do find, however, that due to the constant monotonic parameter of integration of the action integral that we identify as time, the Hamiltonian schema, like the Lagrangian schema, relies on the Newtonian notion of time as an external and equably flowing parameter. Thus despite the variational principle on which the temporal structure of Hamiltonian me-

chanics is built, it is unsurprising that the Newtonian picture of time is standardly attributed to arise from the Hamiltonian schema.

Thus while we find that there are subtle differences to the pictures of time that arise in each of Newtonian, Lagrangian and Hamiltonian mechanics, the Newtonian picture of time remains at the heart of the standard interpretation of all three. The differences that we find between these three formulations of classical mechanics are largely related to the ‘spaces’ in which we describe dynamical behaviour: in the Newtonian case we use a classical spacetime; in the Lagrangian case we use configuration space; and in the Hamiltonian case we use phase space. While each space gives us a varying insight into how Newtonian time can be seen to generate a dynamical reality, we are left with an independent, external, constant monotonic parameter in each case that generates dynamical evolution. To this extent, then, the Newtonian picture of time remains steadfastly attached to the picture of reality that arises from classical mechanics.

It would not be controversial to suggest that the Newtonian picture of time is at the heart of many of the modern metaphysical theses concerning time. The beginning of the 20th century, however, was witness to a major development in the field of mechanics. This development, the advent of relativistic mechanics, changed the way we think about both space and time. The extent to which the metaphysical views built upon the Newtonian picture of time survive these changes is the concern of the next chapter.

Relativistic Constraints for a Naturalistic Metaphysics of Time

In the last chapter we considered the mathematical formalism of three formulations of pre-relativistic classical mechanics with a view to introducing the formal temporal structure we find within more familiar classical physical theories. We discovered there that the Newtonian picture of time that arises from Newtonian mechanics prevails throughout both Lagrangian and Hamiltonian mechanics, despite the characterisation of dynamics varying between these formulations. In this chapter I extend this examination of temporal structure to relativistic mechanics, but here our focus broadens from the more mathematical considerations of the last chapter to include considerations of the orthodox philosophical conceptions of time. While we retain as our main concern the picture of time that arises in physical theory, the purpose of this chapter is to address how the picture of time we find in the theory of relativity comes to bear on the various metaphysical positions in the philosophy of time.

2.1 Introduction

The A-theory of time, stated briefly, proclaims that temporal passage is an objective feature of reality.¹ Implicit in this view is that the temporal instant that embodies this passage, the present, maintains a privileged status over and above the temporal instants that have already ‘passed’ (the past) and that are yet to ‘pass’ (the future). In contrast, the B-theory of time is characterised by its rejection of temporal passage as a real and objective feature of the world. As such, there is no privileged instant and all times from the beginning of the universe to the end of the universe are

¹Of course, there are various A-theoretic views that can be distinguished; more on this shortly.

considered to be equally real according to this view. The division between these opposing temporal theories defines what we will call the *traditional metaphysical debate* on the nature of time. It has been suggested that Einstein's special theory of relativity seriously compromises the viability of various formulations of the A-theory of time²; Minkowski's formulation of the special theory of relativity as a four dimensional spacetime has been instrumental in creating the perception that it provides strong evidence for a B-theory of time. On the other hand, much work has been carried out attempting to show the compatibility of special relativity and A-theories of time³ with a general sentiment emerging that Minkowski spacetime is the wrong sort of entity to definitively adjudicate either way on the traditional debate in the philosophy of time.

This chapter is not an attempt to enter this debate and argue for or against either the A- or B-theory of time; nor is it a concern of this chapter to attempt to argue the consistency of either of these temporal models with classical relativity theory. The purpose of the current analysis is to investigate and outline the constraints, imposed by the temporal structure of classical physical theory⁴, that the traditional debate must heed to remain within the bounds of a naturalistic metaphysics. As one can infer from the introductory remarks above, the special theory of relativity has been conspicuously present in the traditional debate and, therefore, this might make one wonder whether such a project is already *fait accompli*. There are two reasons to be cautious of this presumption. To begin with, existing attempts to answer the question as to why the formal temporal structure of Minkowski spacetime does not preclude the possibility of objective temporal passage (some of which we will meet in §2.4) appear to lack a precise characterisation of the picture of time that arises in special relativity. The initial goal of this chapter is to adopt a *formal characterisation* of time in special relativity (§2.5) with the resulting picture providing a new perspective on why the constraints imposed by special relativity on the traditional debate are not so restrictive as to quash the debate. The second reason is that the temporal structure of general relativity must be considered also if one is to remain within the bounds of a naturalistic metaphysics. The ultimate goal of this chapter,

²See, for instance, Rietdijk (1966), Putnam (1967), Maxwell (1985) and Saunders (2002).

³See, for instance, McCall (1976), Hinchliff (1996), Tooley (2000), Zimmerman (2008) and Savitt (forthcoming).

⁴Other physical theories, especially quantum theory, may impose further constraints on our temporal models but these will not be considered here.

then, is to extend the precise characterisation of time in special relativity to general relativity which, as we will see, imposes much more restrictive constraints.

2.1.1 Background

The traditional metaphysical debate has its origin in an analysis due to McTaggart (1908), who is credited with clearly distinguishing two ways in which we differentiate positions in time (the contents of these positions being events): each position is either past, present or future; or each position is either earlier than or later than some other position. The series of positions running from the past to the present, and then from the present to the future McTaggart labels the A-series. The A-series is characterised by the position of the present; those positions in time earlier than the present are in the past and those later than the present are in the future. The present is in some sense in motion through the A-series as positions in time change from future to present, then from present to past. The series of positions which run from earlier to later independently of the present McTaggart labels the B-series. The B-series is not defined by any position in time labelled as the present. Any position in time is related to any other position in time by either the ‘earlier than’, ‘later than’ or ‘simultaneous with’ relation irrespective of which position might be referred to as the present. These relations are defined irrespective of the position of some present instant and are thus unchanging and objective.

By distinguishing between the ways that we differentiate positions in time, we can construct two types of temporal models. The A-theory of time is the view that an adequate description of temporal reality requires either the A-series alone, or both the A-series and the B-series together. The A-theory is often referred to as a *dynamic* view of time. We will characterise dynamic time here as the claim that we exist in a privileged present that is in some sense ‘flowing’ through successive instants of time.⁵ The present is thus conceived as a privileged element of our reality which demarcates the past from the future in some objective respect. There are two ways that we can understand this privileged present. We can understand the present as ontologically privileged, whereby the notion of flow is envisaged as the existential displacement of the privileged time instant by its successor. Accordingly,

⁵For a nice illustration of some of the various ways this conception of dynamic time can be expressed, see Williams (1951).

each time instant then ‘comes into’ existence as the present instant and then ‘goes out of’ existence as a new time instant becomes the present. We can alternatively understand the present as metaphysically privileged, whereby flow is interpreted as the evolution of some property of ‘presentness’ across consecutive time instants that are ontologically undifferentiated.⁶

The second of the temporal models that we can construct, the B-theory of time, is the view that a complete description of temporal reality can be given by the B-series. The B-theory is often referred to as a *static* or *block universe* model of time. We will take the block universe model to be characterised by two claims: every position in the B-series exists (the past, present and future are equally real), and the A-series is unreal (there is no privileged instant nor any objective flow of time). One can imagine constructing the block by arranging all the positions of the B-series as an unchanging temporal map of the universe and then augmenting this with the spatial dimensions of the universe to create a four dimensional block which contains all the spatial and temporal relations between events. The block universe view forges a strong analogy between the static conception of time and our ordinary conception of space; there is nothing objective about labelling a particular position in space ‘here’ nor claiming the contents of ‘here’ to be more real than the contents of ‘there’, just as there is nothing objective about labelling a particular time ‘now’ whose contents can be thought of as any more real than the contents of any other position within the block. As we will see in §2.2, there exist persuasive arguments that this is more than just a mere analogy.⁷

⁶It can be argued that the metaphysical notion of flow falls afoul of McTaggart’s original analysis of the A- and B-series of time, since a changing property such as presentness requires a separate temporal dimension in which to change, which is thought to be undesirable. We will leave this issue to one side here.

⁷The distinction here between A- and B- theories of time follows that of Dainton (2001, p. 11). The A- and B-theories of time can also be characterised as the ‘tensed’ and ‘tenseless’ theories of time, respectively (as in Le Poidevin (1998), for instance). Under such a construal, the A-theory takes the properties picked out by terms such as past, present and future (known as ‘tenses’) to be real, i.e. to be objective properties of reality. On the other hand, the B-theory denies the reality of tenses. Despite this alternative construal, the core difference between the A- and B-theory remains whether temporal passage is objective or not, as in the above characterisation, and thus ‘tense’ is not taken to be a significant notion here.

2.1.2 Outline of the chapter

This investigation will proceed as follows. I begin in §2.2 by introducing the special theory of relativity, paying particular attention to the geometric formalism and temporal structure of Minkowski spacetime. The dual formulation of time in special relativity as both proper time and coordinate time is also introduced here. As a brief aside to the main issue, I examine in §2.3 an argument from special relativity for the equal reality of the past, present and future. I suggest that this is not a metaphysical constraint that conclusively follows from the formal temporal structure of Minkowski spacetime. In §2.4 I sketch an argument from the literature, which I call the proper time argument, to the effect that neither static nor dynamic views of time are precluded solely by the formal temporal structure of Minkowski spacetime and briefly examine some possible explanations as to why this might be the case. I suggest that the proper time argument crucially turns on the dual formulation of time in special relativity. I set out in §2.5 a precise framework for characterising time and use this framework to formalise the ambiguity with respect to the dual formulation of time in special relativity. I employ this formalism to argue for a more general explanation as to why the temporal structure of Minkowski spacetime precludes neither metaphysical position in the traditional debate. In §2.6 I extend the analysis to time in the general theory of relativity and set out the classical constraints that must be respected by a metaphysical theory of time to remain within the scope of a naturalistic metaphysics.

2.2 Minkowski spacetime

Einstein (1952) developed the special theory of relativity in a 1905 paper, on the shoulders of the pioneering work of FitzGerald, Lorentz and Poincaré. He was motivated in part by the fact that any model of space and time must tell some story about how information concerning distant spatial and temporal regions is made available to a spatiotemporally bound observer. He realised that it is *light signals* that connect us with distant parts of space and time. Einstein used this insight to construct a principled theory of space and time that forms an axiomatic basis for deriving the Lorentz transformations. In 1908, Minkowski (1952) built upon Einstein's special theory of relativity by formally uniting the structure of space and time into a four dimensional object, which we call spacetime: three dimensions representing

space and one dimension representing time. The significance of Minkowski's extension of Einstein's theory is that he formulated the theory in a geometric framework. In this section I introduce the geometric structure of Minkowski spacetime, paying particular attention to the picture of time that arises therein.⁸

Minkowski spacetime can be represented by a geometry $(\mathcal{M}^4, \eta_{\mu\nu})$, which consists of a differentiable, four dimensional manifold, \mathcal{M}^4 , and a flat Lorentzian metric, $\eta_{\mu\nu}$, with signature $(1, 3)$. The manifold is interpreted as representing the set of spacetime points and the metric can be interpreted as the geometric instructions by which these spacetime points are connected. Given a particular point p in \mathcal{M}^4 , and a four-vector dx^μ in the tangent space $T_p\mathcal{M}^4$ at p , we can use the metric to construct the line element, $ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu$, for each spacetime point in \mathcal{M}^4 . The invariance of ds^2 according to the special theory of relativity endows Minkowski spacetime with a conformal structure: we say that dx^μ is a *timelike* vector if $ds^2 > 0$, a *lightlike* vector if $ds^2 = 0$ and a *spacelike* vector if $ds^2 < 0$. The metric thus determines a *lightcone structure* in the tangent space at every point of \mathcal{M}^4 , where the lightlike vectors define the boundary of the cone, the timelike vectors the inside of the cone and the spacelike vectors the outside of the cone.

Since we want to be able to use this formalism to model the behaviour of objects in spacetime, we need to extend these classifications to *curves* in \mathcal{M}^4 . We say that a curve is timelike, lightlike or spacelike if its tangent vector field is characterised as such at every point. We can now interpret timelike curves as the possible spacetime paths of massive particles and lightlike curves as the possible spacetime paths of massless particles (i.e. photons); the actual paths of such objects in spacetime are called *worldlines*. This then gives us a causal structure to Minkowski spacetime: an observer situated at any position in spacetime can divide their surrounding spacetime into a causally contiguous region (the timelike region plus the lightlike boundaries) and causally separated (spacelike) region (see Figure 2.1). The division of spacetime in this way is unique to each spacetime point.

We say that a spacetime \mathcal{M}^4 is *temporally orientable* if there exists a continuous timelike vector field on \mathcal{M}^4 . Minkowski spacetime is temporally orientable. We can then stipulate a temporal orientation to this vector field, simply by picking a future direction, and thus define any timelike or lightlike vector at a point of \mathcal{M}^4 as

⁸The exposition here mostly follows Malament (2007).

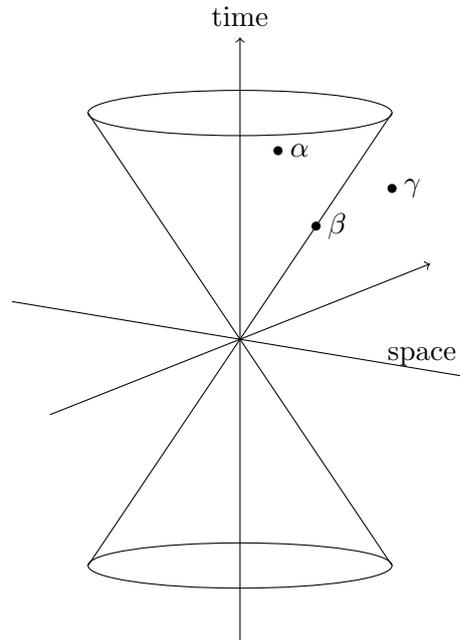


Figure 2.1: The causal structure of Minkowski spacetime for an observer at the origin: α is timelike separated from the observer, β lightlike separated and γ spacelike separated.

future directed or *past directed* with respect to this orientation. As above, a curve is future directed or past directed with respect to this orientation if its tangent vector field is characterised as such at every point. Time is then associated with the parameter employed to parametrise a future directed timelike curve in \mathcal{M}^4 ; such a parametrised curve describes the dynamical behaviour of an object in spacetime (it is only through such a parametrisation that we can begin to speak of ‘time instants’ in special relativity). There are two natural ways that a curve can be parametrised according to an arbitrary observer in spacetime: the curve can be parametrised by the time as measured by a clock moving along the curve in question; or the curve can be parametrised by the time as measured by a clock at rest in some other frame of reference. Let us consider these options more formally.

Given a future directed timelike curve, γ , between spacetime points s_1 and s_2 in \mathcal{M}^4 with tangent field dx^μ , we can define the elapsed time between s_1 and s_2 , τ ,

with which to parametrise γ , as the arc length of the curve:

$$\tau = |\gamma| = \int_{s_1}^{s_2} (\eta_{\mu\nu} dx^\mu dx^\nu)^{\frac{1}{2}} ds. \quad (2.1)$$

The parametrisation of γ by τ is a ‘natural’ parametrisation since the arc length, as a function of the invariant line element, is a frame independent quantity. We thus call τ *proper time* and associate it with the time that a clock will measure along its own (not necessarily inertial) worldline. One can also generate a frame dependent parametrisation of γ : we can employ clocks at rest with respect to some arbitrary reference frame (proper time along a worldline traced out by an object at rest with respect to this reference frame) to define the elapsed time, t , with which to parametrise γ . By employing this method of parametrisation we have, in effect, stipulated an arbitrary coordinatisation of the manifold, with a time coordinate coinciding with proper time in some arbitrary reference frame, with which to describe spacetime dynamics. We thus call t *coordinate time* and associated it with a global time measure corresponding to the fourth coordinate of the spacetime manifold (so long as the reference frame in question is inertial). Since coordinate time is frame dependent, while proper time is frame independent, the latter is taken to have direct physical significance, while the former is not. This dual formulation of time, as proper time and as coordinate time, in Minkowski spacetime will play a crucial role in the discussion below.⁹

The formal relationship between proper time and coordinate time is given by the Lorentz transformations (which are embodied in $\eta_{\mu\nu}$) and is dependent upon the relative velocity of the reference frames from which each time measure is procured. Time intervals (and the relations between them) as measured for a set of events in one reference frame vary from those measured in a second frame moving relative to the first. In addition, the temporal *order* of events at spacelike separation from an observer is also frame dependent; observers in motion relative to one another may record a different temporal order for the very same observed events. It follows that whether two events are *simultaneous* or not is again frame dependent. Thus there is no absolute fact of the matter as to whether two spacelike separated events

⁹The distinction between proper time and coordinate time as formulations of time in the special theory of relativity has also been emphasised by Kroes (1985) and Rovelli (1995) and, more recently, by Savitt (forthcoming).

stand in either the ‘earlier than’, ‘later than’ or ‘simultaneous with’ relation to each other; this relation is dependent upon the observer’s state of motion. Due to this Lorentzian temporal structure, Minkowski spacetime cannot in general be decomposed into distinctly spatial and temporal elements.¹⁰ However, provided that one has stipulated a particular time coordinate coinciding with an inertial timelike trajectory, one is able to generate a *foliation* of the Minkowski manifold consisting of spacelike slices orthogonal to the trajectory and thus constituting a set of simultaneous events.

We now turn our attention to the relationship between the temporal structure of Minkowski spacetime and the traditional metaphysical debate on the nature of time. It is clear from the characterisation just presented that Minkowski spacetime conspicuously constitutes a four dimensional object; indeed, I believe that it is this fact that is at the heart of the conception that Minkowski spacetime *actually is* the block universe. Upon closer analysis, however, the argument from the formal temporal structure of Minkowski spacetime against the dynamic view of time is not quite so straightforward. As mentioned in the introduction to this chapter, particular A-theories of time have been defended against claims that Minkowski spacetime precludes their possibility and, rather than revisit some well trodden ground, we continue in the following manner.

Firstly, and as a brief aside to the main issue, I examine an argument from special relativity due to Putnam (1967) and Rietdijk (1966) for the equal reality of the past, present and future. I suggest in §2.3 that the same cannot be said conclusively of the temporal structure of Minkowski spacetime, however. We return to our main concern again in §2.4 where I present an argument (following Dieks (2006) and Ellis (2007)) that not only illustrates the sort of analysis that one might provide in support of a dynamic view of time but also starts us in the right direction to providing in §2.5 a more general argument as to why the formal temporal structure of Minkowski spacetime alone precludes neither static nor dynamic views of time. This is all, of course, with a view to describing in §2.6 the constraints prescribed by the general theory of relativity that must be recognised by a metaphysical theory of time to fall within the scope of a naturalistic metaphysics.

¹⁰In contrast, recall that the ordinary Euclidean metric imposed on a four dimensional manifold results in a Newtonian spacetime in which space and time can be globally separated as distinct elements of the manifold.

2.3 Spacetime and reality

Some dynamic views of time make claims about whether or not some or all of the events located in the past, present and future are equally real. Putnam (1967), in his *Time and Physical Geometry*, presents an argument from special relativity against a dynamic view of time called *presentism*; Putnam argues for the equal reality of the past, present and future, which presentists deny.¹¹ In the preliminary exposition of the argument below I follow Putnam's terminology.

Putnam's project is to examine the claim that all and only things that exist now are real (which can be taken as a statement of presentism). He begins by making some assumptions about what it means for something to be real: the property of 'reality' is assumed to be parasitic on a coordinate independent, transitive relation which Putnam labels R . To illustrate this relation as clearly as possible consider two observers in spacetime, Amy and Ben. If we consider Amy as being in the present, we can take Putnam to be deeming as real all and only those 'things' standing in the R relation to Amy. He makes the further assumption that if this is the case, and Ben is also real, then all and only those 'things' standing in the R relation to Ben are real. It is clear that Putnam's relation R and its significance for reality is intended to capture a conception of pre-relativistic classical simultaneity. It is the incompatibility between this conception of reality and the structure of Minkowski spacetime that is the place that Putnam wishes to drive his wedge against presentism. Let us walk through Putnam's argument to demonstrate this.

If the relation R were merely the relation of simultaneity in a pre-relativistic classical spacetime, then the above assumptions about reality would lead one to accept the claim that all and only things that exist now are real, i.e that the past and future are not real. However, when we move to Minkowski spacetime things are not as straightforward. Recall at this point the physical consequences of the Lorentzian metric, $\eta_{\mu\nu}$. If Ben is in motion relative to Amy, the events in spacetime simultaneous with Ben will not coincide with the events in spacetime simultaneous with Amy. In fact, some of the events simultaneous with Ben will be in regions of spacetime Amy considers her future, and some will be in regions of spacetime Amy considers her past. Thus if we were to conceive of the relation R as the relativistic

¹¹Rietdijk (1966) presents a similar argument, but the clarity of Putnam's analysis will be instrumental in what follows below.

relation of simultaneity (that is, “ x is simultaneous with y in the coordinate system of x ” (Putnam, 1967, p. 242)), we would be led to the conclusion that the events that lie in Amy’s past and future have the same claim to reality as the events she considers her present. However, according to Putnam we are not at liberty to conceive of the relation R as the relativistic relation of simultaneity since the former must be transitive and the latter is not, and it is the transitivity of R that is doing the work of ensuring that Ben and Amy agree about what is real.¹²

To overcome this, Putnam stops short of equating the relation R with the relativistic relation of simultaneity and instead claims that the relation R merely holds between ‘things’ which are on the same plane of simultaneity. This then, according to Putnam, coherently ensures the equal reality of the past, present and future: those events that are simultaneous with Amy stand in the R relation to Amy and are therefore real; Ben is simultaneous with Amy; those events simultaneous with Ben (including events in Amy’s past and future) stand in the R relation to Ben and are therefore also real; thus the past, present and future have equal claim to reality. The use of the relation R by Putnam now becomes a bit clearer. To run the argument without the R relation would require one to tie reality directly to simultaneity. Since simultaneity is not transitive, one would have to concede that reality were transitive for the argument to be successful. However, or so one could argue, reality simply cannot be a two-place relation: something is either real or not - it cannot be the case that something is real *for an observer*. Thus Putnam’s solution is to have the transitive relation R piggyback on simultaneity and reality in turn is then parasitic on the relation R .

The crucial point to this argument is the nature of simultaneity; it is upon the back of the relativistic simultaneity relation that both the R relation and reality ride. Recall, though, what we have already said to be the case in special relativity: there is no objective fact of the matter as to which events are simultaneous with a particular spacetime event. This is because observers at the same place in spacetime might be in relative motion to one another and therefore make different judgments about the temporal ordering of spacelike separated events.

As it happens, the set of events that we might class as simultaneous with any

¹²As pointed out by an anonymous referee, the above quoted relativistic relation of simultaneity is *not* a single relation that fails to be transitive but is in fact an equivalence class of such relations, the utilisation of any two of which enable an argument such as this one of Putnam’s.

particular observer can only be classed as such by convention, as Einstein notoriously noted. Since all the events that are spacelike separated from a particular observer are epistemically inaccessible to that observer, we must stipulate by convention which set of events are to be the ones which constitute a plane of simultaneity; any such choice would be consistent with the local phenomenology of that observer. Because of this, any way one chooses to relate the reality of spacelike separated events to an observer must come prepackaged with metaphysical assumptions about the reality of the various parts of spacetime. This jeopardises the cogency of Putnam's result; let us see how explicitly.

When Putnam refers to the events in spacetime simultaneous with Ben, for instance, it has not been explicitly stated what this might mean. It might be implicit here that this is the set of events on a spacelike hypersurface orthogonal to Ben's worldline. Likewise, it has not been explicitly stated what is meant by the events Amy considers to be her future and past. One would expect that the events Amy considers to be in her future are those events in her future lightcone, and the events Amy considers to be in her past are those in her past lightcone. However, that Amy can label events at spacelike separation from her as past or future would suggest that Putnam has something more in mind, perhaps events in the past or future of a spacelike hypersurface orthogonal to Amy's worldline. By being explicit about what is meant in Putnam's discourse, it no longer seems at all clear how the relation R , which 'merely' holds between 'things' that are on the same plane of simultaneity, can achieve all that it is claimed to achieve according to Putnam.

With respect to the conventionality of simultaneity, Malament (1977) presents an argument that the standard simultaneity relation (namely, the convention utilised in Putnam's argument) is uniquely definable from the geometry of Minkowski spacetime. If we imagine a single worldline γ through $(\mathcal{M}^4, \eta_{\mu\nu})$, and two points p and q on γ , the intersection of the past lightcone of the later point and the future lightcone of the earlier point will define uniquely a spacelike hypersurface that coincides with a notion of four dimensional orthogonality commensurate with Einstein's convention for simultaneity. Thus it appears as though the conformal structure of Minkowski spacetime eliminates the need for a convention. In response to this, Brown (2005, §6.3.1) suggests that there is something not quite right with this argument. Malament's treatment of simultaneity assumes that the geometric structure of Minkowski spacetime is dissociated from the spatiotemporal objects it describes. The argument

employs the conformal structure of Minkowski spacetime and just a single worldline, γ , but, and this is the point that Brown makes, it is not entirely clear that a spacetime consisting of a single worldline has such a conformal structure. Without delving into Brown's constructive project too deeply, he maintains that the geometry we attribute to spacetime is merely a convenient representation of physical laws governing the behaviour of matter and therefore is not the causal explanation of this behaviour. Thus the geometric structure of a spacetime comprised of a single worldline is not necessarily that of Minkowski spacetime. To make matters worse, Janis (1983) argues that the addition of any extra content to Malament's spacetime would compromise the uniqueness, and thus the non-conventionality, of any subsequent simultaneity relation. At the very least we should be wary of Putnam's use of this simultaneity relation.

Dickson (1998, §8.1.2) (following Stein (1968)) offers an alternative appraisal of Putnam's argument. According to Dickson, Putnam's conclusion, that the past, present and future are equally real, is framed in language that simply does not make sense in a relativistic context; the relativity of simultaneity ensures that the notions of absolute past, present and future are nonapplicable here. In Dickson's words:

Putnam considers the doctrine that all and only those events that are *now* are real. He concludes, on the basis of special relativity, that this doctrine is false, and takes himself to have shown thereby that 'the future' is real. This form of argument is apparently not very compelling. (Dickson, 1998, p. 170)

Yet another suggestion that has been made in response to Putnam's argument, due to Sklar (1977, p. 277), is that we can simply bite the bullet and take reality to be a transitive relation. Thus, along with the relativity of simultaneity, special relativity would suggest that reality is relative also. The essence of the argument is that the theory of relativity is counterintuitive to begin with, so why not the reality relation as well? The relativity of reality would suggest that Amy could claim a set of spacetime events as real, and Ben could claim a different set of spacetime events as real and there would be no fact of the matter about which observation (if either) was an accurate representation of the world. Although, such a fragmented ontology may be undesirable for other reasons.

A complete analysis of the cogency of Putnam's argument will not be a concern of this project; indeed, there are metaphysical arguments that can be made in defence of Putnam. What I hope to have suggested here, however, is that the

issue concerning whether the equal reality of the past, present and future follows straightforwardly as a constraint on the metaphysical nature of time from the formal temporal structure of Minkowski spacetime is more contentious than Putnam would have us believe.

2.4 Objective temporal passage

The most pressing concern for an A-theorist when presented with Minkowski spacetime is the question of how to endow the manifold with an objective temporal passage. Since temporal passage invariably involves change, for Minkowski spacetime to include temporal passage as an objective element something within the manifold would have to undergo some sort of ‘change’. The most obvious candidate for this “something” is an objective ‘now’: a hyperplane of simultaneity within spacetime which privileges a particular time instant and which embodies the passage of time. The problem at this point for the A-theorist is that no such hyperplane of simultaneity is privileged as such; due to the relativity of simultaneity, many hyperplanes of simultaneity can be specified depending on the relative motion of the observer and none of these can claim any special status as being a privileged time instant. Thus, it seems as if there is no scope for an objective ‘now’ and thus no scope for objective temporal passage.

Not all is lost for the dynamic view of time though. While this argument does provide some important restrictions on the form that an objective temporal passage can take, it does not show that objective temporal passage is incompatible with the formal temporal structure of Minkowski spacetime. There is no objective hyperplane of simultaneity in Minkowski spacetime and thus no objective *global* ‘now’.¹³ However, a global ‘now’ is not the only candidate for the basis of temporal passage. While an integral element of the special theory of relativity is that there is no absolute fact of the matter about global temporal orderings, there are some facets of Minkowski spacetime that are absolute. Recall (Figure 2.1) that the conformal structure of Minkowski spacetime separates the manifold into timelike separated events (inside the light cone) and spacelike separated events (outside the light cone). Observers at the same position in spacetime but in motion relative to one another will define their hyperplane of simultaneity and their local direction of

¹³Unless, of course, one adds some extra structure.

time skewed with respect to one another, but the conformal structure of Minkowski spacetime is inherent in the geometry; they will agree on which regions of spacetime are timelike separated and which regions of spacetime are spacelike separated.

We encountered the claim above that according to relativity theory there is no objective fact of the matter as to whether two spacelike separated events stand in either the ‘earlier than’, ‘later than’ or ‘simultaneous with’ relation to each other. In contrast, the causal structure of Minkowski spacetime permits that for future directed timelike curves there *is* an objective fact of the matter as to which events are past and which events are future. This temporal ordering of events is only local (i.e. applicable to a single point on a worldline) since observers with varied relative motions will disagree on the ordering of spacelike separated events. One can then imagine any single spacetime point on a future directed worldline as a candidate for an objective *local* ‘now’. Minkowski spacetime would then contain many such local objective ‘nows’, each associated with a single worldline. The formal geometric structure of Minkowski spacetime then does *not* preclude the possibility of an objective local ‘now’ (though it certainly does limit the scope of such a ‘now’) and therefore does not preclude out of hand this particular form of objective temporal passage. Let us call this argument the *proper time argument*.

It is far from obvious that the metaphysical notion of dynamic time that arises from the proper time argument is indeed a viable metaphysical position. The A-theorist who wishes to develop such a view requires an explanation of exactly how consistency can be maintained between the dynamic local nows, and moreover, in such a way that we might recover something resembling our normal experience of spatiotemporal events. I contend, however, that if the resulting picture of time is apt to be rejected as unfavourable, this would evidently not be as a result of the formal geometric structure of Minkowski spacetime. These issues aside, it is interesting to ask then why, as a result of the proper time argument, we find ourselves unable to undermine either static or dynamic views of time with solely the temporal structure of Minkowski spacetime. Let us briefly consider two possible explanations.

Dieks (1991) makes the suggestion that it is the universality of physical theories that prevents them from including specific metaphysical commitments concerning the flow of time. According to Dieks, physical theories are concerned with the task of giving descriptions of universal laws, valid at all times and places and this can only be achieved if all times and places are treated on an equal footing; there are no

times or places to which the laws of physics must be anchored. This is the source of the purely relational nature of physical laws. An absolute and global difference between past and future, for instance, simply does not and cannot exist in a physical theory. The specific properties of events are disregarded in physical theories and only what is common to all processes of a particular kind is retained. As Dieks puts it, “the laws of physics by themselves cannot reveal what time it is” (1991, p. 258). Thus according to this view, the ‘now’ of experience may indeed reflect something objectively real but we should not expect it to play a role in our physical theories, including Minkowski spacetime, and thus our physical theories should not be able to provide evidence for any particular metaphysical view of time.¹⁴

Another possible explanation as to why the formal temporal structure of Minkowski spacetime might not preclude either static or dynamic views concerns the underdetermination of metaphysics by physics. One could argue that we should never expect conclusive determination of the metaphysics of our best physical theories, in the same way that we do not expect complete determination of our physical theories from observation. This does not, however, seem to provide the explanation we are looking for. The claim against which the proper time argument is levelled is that formal geometric structure of Minkowski spacetime precludes the possibility of a dynamic view of time. Therefore, just in the same way that it is possible for an observation to falsify a physical theory, so it is possible that a physical theory falsify one or more of our metaphysical beliefs. So although we should not expect any particular physical theory to provide definitive evidence for any particular metaphysical view, it seems reasonable to expect that a physical theory might eliminate the possibility of a certain metaphysical thesis.¹⁵

The more compelling explanation, which has already been alluded to in the previous section, is that the proper time argument turns on the ambiguity we find in the picture of time that arises in the special theory of relativity. We saw two ways in which time is formulated in special relativity: the first is as a time measure along

¹⁴This sentiment is echoed by McCall (2001, p. 144): “Although not endorsing time flow, modern physics does not rule it out as a logical impossibility. Its attitude is empirical rather than logical; time flow may exist, but (i) there is no hard experimental evidence for it, and (ii) it plays no role in physical theory.” Zimmerman (2008, p. 220) also makes a similar point.

¹⁵A lengthy discussion of this issue has played out in the literature with respect to the metaphysical underdetermination of quantum field theory; see van Fraassen (1991), French (1998, 1989), Ladyman (1998) and French and Ladyman (2003).

an individual worldline, proper time; and the second is as a time measure associated with a coordinatisation of the manifold, coordinate time. Due to the conformal structure of Minkowski spacetime, there are restrictions on how a particular manifold can be foliated. However, given these restrictions there still remains an infinite number of ways to coordinatise the manifold. Thus we are not left any option for stipulating a global objective ‘now’. Although objective temporal passage cannot correspond consistently with some objective time coordinate of the manifold, we *are* able to imagine that objective temporal passage corresponds with the incremental evolution along an object’s worldline, or the proper time in some reference frame (namely, the reference frame that contains the object in question). This variable characterisation of time in the special theory of relativity thus gives us a good clue as to why an argument can be made *against* the possibility of dynamic time in Minkowski spacetime, in the first place, and why Minkowski spacetime does not formally *preclude* either metaphysical position in the traditional debate, thereafter.

Having said this, however, it is not obvious that the opposing views of the traditional debate emerge from these considerations unscathed, either. For instance, it might seem that relying on the local proper time along a worldline to locate objective temporal passage indeed results in quite a significant modification of the metaphysical position that the A-theorist originally intended. Thus depending upon which features of dynamic time the A-theorist thinks essential, the possibility arises that the metaphysical theory resulting from the above considerations does not do justice to dynamic time. We must keep in mind at this point, however, that rejecting the logical space circumscribed by the proper time argument, i.e. rejecting the authority of contemporary physical theory, is tantamount to rejecting a naturalistic metaphysics.

The discussion of this section has conspicuously lacked the formal machinery with which we introduced the geometry of special relativity in §2.2. Let us turn then to the promised formal characterisation of the temporal structure of Minkowski spacetime; doing so will serve to illustrate precisely why the proper time argument functions as it does.

2.5 Characterising time

In both his (1995) and his (2004), Rovelli sets about characterising the various roles that the concept of time plays in different scientific theories (recall the discussion from the Introduction).¹⁶ The terminological project associated with this analysis is complicated by the multitude of features that are attributed to the concept of time in natural language. Not often are the entirety of these features found bundled together in the formal structure we identify as time in a physical theory. I will adapt Rovelli's formalism here to provide a more precise grounding to the explanation above with respect to the temporal structure of Minkowski spacetime. Let us begin by considering time as it is often characterised, as a variable t which parametrises the real line \mathbb{R} .

The real line can be described by the following structure: a manifold, \mathcal{M}^1 , consisting of a set of objects (which in this case is simply all the real numbers) with a one dimensional topology and a differential structure; an ordering, $<$, which sequences the members of the set within the topological structure; a metric, g , which ensures that the distance between any two members of the set is meaningfully measurable; and an origin, φ , which fixes a preferred member of the set. Let us represent this as $\mathbb{R} : \{\mathcal{M}^1, <, g, \varphi\}$ (Rovelli, 1995, p. 83). It is clear to see that this structure maps into the features we ordinarily associate with the notion of time; the set of objects represent the instants of time, the ordering represents the sequential structure of the instants, the metric represents a measure of temporal duration and the preferred fixed time instant is the present. We should note, however, that this short list of attributes represented by the real line is not a consequence of any particular physical theory. If we consider the picture of time in Newtonian mechanics, for instance, recall from §1.2 that there is no preferred fixed point in the theory that is necessarily labelled as the present. This is not to say that the characterisation of time as the real line is incompatible with Newtonian time; on the contrary, time characterised by the real line is quite consistent with the temporal structure of Newtonian theory. Let us represent the structure of Newtonian time as $N : \{\mathcal{M}^1, <, g\}$ and represent that it is consistent with a richer structure by $N : \{\mathcal{M}^1, <, g \mid \varphi\}$.

¹⁶As well as Rovelli, the different features of time have also been discussed with respect to the special theory of relativity by Kroes (1985) and with respect to both special and general relativity by Callender (2006).

As was introduced in §2.2, the characterisation of time in special relativity is not so straightforward. We saw that the dynamical behaviour of objects in spacetime is described by future directed timelike curves in a four dimensional geometry, $(\mathcal{M}^4, \eta_{\mu\nu})$, and that the notion of time is associated with the parametrisation of such curves. The significant feature of time in special relativity that sets it apart from Newtonian time is that, for all p in \mathcal{M}^4 , a whole family of future directed timelike curves through p provide a multitude of candidate structures with which time might be identified; the conformal structure of the Minkowski geometry simply does not permit a unique global one dimensional time to be defined in terms of the geometric structure of \mathcal{M}^4 . In other words, it is not possible to define, in terms of the geometric structure of spacetime, a global ordering of all the spacetime points in \mathcal{M}^4 ; we can only define a partial ordering on the set of spacetime points, $<'$.¹⁷ There are, however, two avenues open to us for reinstating a total ordering to a set of spacetime points in \mathcal{M}^4 which correspond to characterising time as coordinate time and proper time, respectively. Let us consider coordinate time first.

While the structure of Minkowski spacetime may not permit a unique global one dimensional time to be directly *definable* from the geometric structure of \mathcal{M}^4 , we are at liberty to *impose* such a structure on the set of all spacetime points. We can simply choose an arbitrary reference frame and take the time as measured by clocks at rest in that frame to provide a unique foliation of the manifold. Of course, a global time measure of this sort is just coordinate time and the unique foliation of \mathcal{M}^4 into hyperplanes of simultaneity does indeed yield a one dimensional set of time instants (the global hyperplanes), \mathcal{M}^1 , with a total ordering, $<$. A caveat arises at this point, however, when one considers that there is an uncountably infinite number of ways that one can choose such a coordinatisation of the manifold. For every inertial future directed timelike curve through some $p \in \mathcal{M}^4$ there is a corresponding foliation of the manifold. Thus there is an infinite number of ways that one might measure the temporal duration between any particular pair of events, corresponding to an

¹⁷A *total order* on a set S is defined by a binary relation (\leq) with the following properties:

- (i) $\forall x \in S, x \leq x$,
- (ii) $\forall x, y \in S, x \leq y \ \& \ y \leq x \Rightarrow x = y$,
- (iii) $\forall x, y, z \in S, x \leq y \ \& \ y \leq z \Rightarrow x \leq z$, and
- (iv) $\forall x, y \in S, x \leq y$ or $y \leq x$.

A *partial order* on a set is a binary relation that satisfies (i)-(iii) but not (iv).

infinite number of reference frames, and thus any such measurement in an arbitrary coordinate system is physically meaningless. The characterisation of time in the special theory of relativity as coordinate time thus lacks global *metricity* (i.e. a unique global measure of time). Thus for any reference frame F , we can represent coordinate time in special relativity as $C_S(F) : \{\mathcal{M}^1, <\}_F$.

A second methodology that we can adopt to find a total ordering of a set of spacetime points in \mathcal{M}^4 is to restrict ourselves to a subset of points in the manifold. Rather than search for a unique *global* one dimensional time, we can instead make use of the linear structure of a single future directed timelike curve to provide a *local* measure of time. Of course, a time measure of this sort is just proper time and the local parametrisation of such a curve yields a one dimensional set of time instants, \mathcal{M}^1 , with a total ordering, $<$. Since proper time is an invariant time measure, the associated parametrisation of a particular worldline is observer independent and thus is a physically meaningful time measure (of temporal durations along the curve only), i.e. proper time is locally metrical. Thus for any timelike curve γ , we can represent proper time in special relativity as $P_S(\gamma) : \{\mathcal{M}^1, <, \eta_{\mu\nu}\}_\gamma$. In addition, since proper time is only defined locally, fixing a preferred time instant amounts to privileging merely a single spacetime point rather than some global hyperplane. Thus a preferred fixed time instant is consistent with the structure of proper time, $P_S(\gamma) : \{\mathcal{M}^1, <, \eta_{\mu\nu} \mid \varphi\}_\gamma$.

As a final element to analysing the proper time argument, let us attempt an equally precise construal of dynamic time. We are taking the dynamic view of time here as the claim that we exist in a privileged present that is in some sense ‘flowing’ through successive instants of time. Let us consider which of the above attributes might best fit with this notion of time. Dynamic time is certainly linear, has a well defined order (directed towards the future) and fixes a preferred time instant (the present). Inherent in the idea of ‘flow’ is a notion of continuity that is meaningful only when there exists a measure across the flowing time instants, i.e. dynamic time requires a definite metric. Thus it seems as though dynamic time can be construed as having the structure of the real line as above, $D = \mathbb{R} : \{\mathcal{M}^1, <, g, \varphi\}$ (which is hardly a surprise). Let us now reconsider the proper time argument in light of these considerations.

The charge was made against the A-theorist that there can be no objective temporal passage in Minkowski spacetime because there is no scope for an objective

hyperplane of simultaneity. This amounts to a claim that not only is there no preferred time instant in special relativity, but a preferred time instant is incompatible with the temporal structure of special relativity. It is clear that this argument aims to characterise time in special relativity as coordinate time and, in light of the above analysis, $C_S(F) \neq D$; not only is $C_S(F)$ incompatible with a preferred time instant, $C_S(F)$ is incompatible with any global and physically meaningful definition of a metric. If the structure of coordinate time were the only formulation of time in special relativity then, to stay within the bounds of a naturalistic metaphysics, dynamic time as we have presented it here would need to be reconsidered as a metaphysical position.

However, we know that time can be construed in special relativity in terms of the structure $P_S(\gamma) : \{\mathcal{M}^1, <, \eta_{\mu\nu} \mid \varphi\}_\gamma$. By formalising the temporal structure of both Minkowski spacetime and the dynamic view of time in this way, we can see immediately that $P_S(\gamma)$ is completely consistent with $D : \{\mathcal{M}^1, <, g, \varphi\}$ (given that $\eta_{\mu\nu}$ and g are both ‘flat’ metrics). Thus the dynamic view of time is *not* precluded by the formal temporal structure of Minkowski spacetime. This is then the more compelling explanation as to why the proper time argument functions as it does: the picture of proper time that arises in special relativity ensures that the dynamic view of time is compatible with the temporal structure of Minkowski spacetime due to the correspondence between the characterisations of time that each of them yield. The constraints imposed by the temporal structure of Minkowski spacetime on a naturalistic metaphysics are thus not so restrictive as to force an A-theorist into a major rethink of her position (or a B-theorist, either, for that matter).

This result is the first of the goals I set out at the beginning of this chapter: the precise characterisation of the features of time in special relativity illustrates more clearly why the formal constraints imposed by special relativity on the traditional debate are not so restrictive as to quash the debate. In the next section I wish to address the prime goal of this chapter: to show that the general theory of relativity provides much sterner restrictions on a naturalistic metaphysics of time. The argument which leads us to these restrictions hinges on an additional feature of time that we find within general relativity but not within special relativity.

2.6 The traditional debate constrained

Our description in §2.2 of the dynamical behaviour of objects in spacetime according to the special theory of relativity is in terms of curves through \mathcal{M}^4 ; insofar as this is the case, we are treating spatiotemporal objects as point particles. To provide a more general description of dynamical behaviour in spacetime, we can extend our formalism with the addition of *matter fields*. A matter field is represented by a smooth tensor field, $T_{\mu\nu}$, on \mathcal{M}^4 and is assumed to satisfy field equations relating $T_{\mu\nu}$ and the metric. A crucial element to recovering the correspondence between future directed timelike curves on \mathcal{M}^4 and worldlines of massive particles in spacetime in the special theory of relativity is the latent assumption that the background spacetime structure, $(\mathcal{M}^4, \eta_{\mu\nu})$, remains fixed *independently* of the $T_{\mu\nu}$ that live on \mathcal{M}^4 .¹⁸ We will call time *independent* when the metric defining time is independent of the matter and energy distribution in the manifold and retain the notation $\eta_{\mu\nu}$ for an independent metric. It is the fact that $T_{\mu\nu}$ is independent of the background spacetime in special relativity that allows us to describe the dynamical behaviour of matter in spacetime in terms of evolution in a time parameter with metric properties (proper time). In contrast when a dependency exists between $T_{\mu\nu}$ and the metric the evolution of the system *defines* proper time and not *vice versa*. General relativity is characterised by such a dependency and we denote the dependent metric $g_{\mu\nu}$.

The geometric structure of general relativity is much the same as the structure we introduced in §2.2 for special relativity: we have a geometry $(\mathcal{M}^4, g_{\mu\nu})$ and we define proper time as before (2.1). The dependency between $T_{\mu\nu}$ and $g_{\mu\nu}$ is given by the Einstein field equations,

$$G_{\mu\nu}(g_{\mu\nu}) = 8\pi T_{\mu\nu}, \quad (2.2)$$

that define an explicit relation between the matter/energy content of spacetime, represented by the stress-energy tensor $T_{\mu\nu}$, and the curvature of the spacetime manifold, represented by the Einstein tensor $G_{\mu\nu}$ (which is a function of the metric, $g_{\mu\nu}$). Due to this relation, the metric, of which proper time is a function, is a dynamical entity that is at each point in spacetime directly dependent upon the

¹⁸Malament (2007, p. 242). We can think of an independent $T_{\mu\nu}$ in the special theory of relativity as representing “test particles” in spacetime.

matter/energy density at that point.¹⁹

The picture of time that arises in general relativity holds similarities with the picture that arises in special relativity; although there are some important differences. Coordinate time can be thought of as an arbitrary foliation of the manifold, each which gives a unique slicing of four dimensional spacetime into a sequence of three dimensional configurations. The linear substructure determined by the foliation yields a total ordering of the slices. However, since the metric is a pointwise function of the matter/energy density of spacetime, it is no longer a flat metric as is the case in special relativity. There is thus no unique notion of parallel transport in curved spacetime and hence there is no way to compare velocities at different points in the manifold unambiguously. Moreover, the parametrisation of the time slices defined by coordinate time in general relativity can be arbitrarily rescaled, which forbids any notion of meaningfully measuring time intervals between pairs of events. This compounds the arbitrariness of such a coordinatisation of the manifold as a temporal measure and so destroys any notion of metricity. Thus for any reference frame F , we represent coordinate time in general relativity as $C_G(F) : \{\mathcal{M}^1, <\}$. Again, coordinate time is merely an imposition of an arbitrary variable determining time evolution, and because general relativity is foliation invariant, coordinate time has no physical significance.

Proper time in general relativity, on the other hand, is defined exactly as in special relativity (2.1), except that in general relativity it is determined by a dependent metric, $g_{\mu\nu}$, as above. The local parametrisation of a general relativistic worldline in terms of proper time again yields a one dimensional set of time instants, \mathcal{M}^1 , with a total ordering, $<$. Thus for any timelike curve γ , we represent proper time in general relativity with the structure $P_G(\gamma) : \{\mathcal{M}^1, <, g_{\mu\nu}\}_\gamma$.

If we now consider the structure of dynamic time, D , we can see immediately that similar arguments to those above could be constructed claiming dynamic time to be inconsistent with the structure of general relativity *if* time were characterised simply by $C_G(F)$: coordinate time in general relativity has no physically meaningful metric properties. We know, however, not to be persuaded by such argumentation. The case for the consistency of dynamic time with proper time in general relativity, on the other hand, is not so clear cut. We can compare proper time in special relativity,

¹⁹More accurately, there is a *mutual dependency* between the stress-energy tensor and the metric at each point of the manifold.

$P_S(\gamma) : \{\mathcal{M}^1, <, \eta_{\mu\nu}\}_\gamma$, to proper time in general relativity, $P_G(\gamma) : \{\mathcal{M}^1, <, g_{\mu\nu}\}_\gamma$, and see that the only difference between the two is the dependency of the metric. The proper time argument of §2.4 demonstrated that the possibility of locally privileging a temporal instant in a special relativistic spacetime with an independent metric is not prohibited by the formal temporal structure therein. Whether the same can be said for a general relativistic spacetime with a metric that is a pointwise function of the four dimensional matter/energy distribution remains to be shown. Exploring this possibility will lead us to the constraints that classical physics imposes on the traditional metaphysical debate.

For dynamic time to be consistent with a physical theory there must be a characterisation of time therein that allows us to privilege a present moment that flows objectively. Recall (§2.1.1) that flow can be construed in two different ways depending upon whether we understand the privileged present ontologically or metaphysically. In the special theory of relativity proper time is determined by a fixed background metric structure. The rate of flow of time along a worldline, being determined by the metric, is then not a function of any part of spacetime but the immediate local neighbourhood of the ‘privileged’ instant on the worldline in question; the local flow is determined locally. In this respect, special relativity formally precludes neither an ontologically privileged present nor a metaphysically privileged present, since the flow of time along a worldline does not force us to make an ontological commitment to any part of spacetime but the fixed background structure at a particular spacetime point.

Turning our attention to general relativity, there are two considerations that seem to pull us in opposing directions. The first consideration is that the proper time between two spacetime points on a worldline in general relativity is, just as in special relativity, determined by the metric on the spacetime segment between them, and this is related *locally* (i.e. pointwise) to $T_{\mu\nu}$ via the Einstein field equations. In addition, since it is a principle of general relativity that, for any point of spacetime, we can find a coordinate system in which the metric locally takes the form of the Minkowski metric, there does not appear to be any grounds for a difference between the local properties of time from special relativity to general relativity.

A second consideration, however, suggests that there *is* something global about the ontology of general relativity. Westman and Sonego (2009) argue that in a generally invariant theory like general relativity, it is untenable to endow a coor-

dinatisation of the manifold, x^μ , with operational significance (i.e. as referring to readings on rulers and clocks) since this leads to the underdetermination of Einstein's field equations. Rather, the x^μ must be interpreted merely as mathematical parameters.²⁰ This amounts to the claim that \mathcal{M}^4 cannot represent something empirically accessible in general relativity. What is empirically accessible, according to Westman and Sonego (2009), is the coincident values of different measurable physical quantities (field values) that motivate a refined notion of an event, which they label a "point-coincidence". The set of all point-coincidences, which possesses a natural manifold structure, denoted \mathcal{E} , turns out to be a natural representation of the totality of physical events (i.e. spacetime). In such a representation \mathcal{M}^4 plays no empirical role and only the mutual relationships of the configurations of various fields are physically relevant. Thus, the suggestion is that general relativity must be interpreted as having a kind of *relational* ontology.

It is hard, then, to envisage an ontologically privileged present in a general relativistic spacetime given the portent here that general relativity is predicated upon a coordinate invariant notion of ontology. The exclusivity of the reality of a locally defined present time instant (as required by an ontologically privileged present) seems to be compromised by the relational nature of the ontology of general relativity. Thus one might struggle to justify a metaphysical theory of classical time, which remains within the scope of a naturalistic metaphysics, and interprets flow in terms of the existential displacement of a privileged time instant by its successor.

To remain within these constraints, the traditional debate must proceed in the following manner: if one wanted to maintain that there is an objectively flowing privileged time instant, then one must understand this instant to be metaphysically privileged, whereby flow is interpreted as the evolution of the property 'presentness' across consecutive time instants that are ontologically undifferentiated.

What is far from obvious is whether this picture yields a nontrivial metaphysical theory of time; for instance, in what meaningful sense is this conception of the present 'privileged' or 'objective', especially if we are simply positing a preferred temporal instant with this property to allow us to maintain that we occupy an A-theoretic reality? Whether or not there remains logical space for an A-theory of time within these constraints depends upon the way in which the A-theorist wishes to

²⁰As Westman and Sonego (2009, p. 1594) point out, this does not mean that charts on a manifold are arbitrary; rather, the *manifold points* themselves lack operational significance.

refine the notion of the privileged present.²¹ As an integral element to any ensuing analysis here, I wish to point out in a sceptical spirit that the dynamic view of time seems to be beset by the imprecise and obscure nature of notions such as ‘privileged’, ‘objective’ and ‘flow’ and it is not entirely clear that these terms are conducive to rigorous definition in this context.²² The implication, then, is that the A-theorist who respects naturalistic metaphysics owes us an account of the dynamic view of time that avoids the triviality of merely stipulating a spacetime point as objectively metaphysically distinguished.

There is a further caveat that jeopardises the viability of a dynamic view of time.²³ Even if we consider that each individual worldline in spacetime is a vehicle of objective flow, to ensure that every such worldline yields a totally ordered linear subset of the manifold we require the existence of a spacelike hypersurface $\Sigma \subset \mathcal{M}^4$ with the property that every inextendible timelike curve in \mathcal{M}^4 intersects Σ exactly once. We call Σ a *Cauchy surface* and note that it follows from this condition that Σ is a three dimensional spacelike submanifold of \mathcal{M}^4 . A geometry $(\mathcal{M}^4, g_{\mu\nu})$ that admits the existence of a Cauchy surface is said to be *globally hyperbolic*. If $(\mathcal{M}^4, g_{\mu\nu})$ is globally hyperbolic then \mathcal{M}^4 is diffeomorphic to a manifold of the form $\Sigma \times \mathbb{R}$ (where we take Σ here to represent a diffeomorphism equivalence class of three dimensional Cauchy surfaces) (Geroch, 1970). Thus a necessary condition for the possibility of dynamic time is the requirement that our reality be represented by a manifold \mathcal{M}^4 that can be foliated by Cauchy surfaces.²⁴ A problem arises here for the dynamic view of time since only a subset of the solutions to Einstein’s field equations (2.2) have this property; Gödel’s (1949) infamous and eponymous spacetime solution, which contains closed timelike curves, is just one example of a solution to the field equations that is not globally hyperbolic.²⁵

²¹Zimmerman (forthcoming) sets out a comprehensive defence of an A-theory of time that fits explicitly within these constraints.

²²Though see Price (forthcoming) for a recent (and not very sympathetic) analysis of flow.

²³The formalism here follows Belot (2007)

²⁴Of course, one could argue that a total ordering of temporal instants is not essential to the dynamic view of time, i.e. dynamic time might be better characterised by the structure $D' : (\mathcal{M}^1, <', g, \varphi)$, with a partial ordering $<'$. Global hyperbolicity would not be a necessary condition for the possibility of a dynamic time represented by D' .

²⁵On the other hand, just because a physical theory admits the *possibility* of a solution to the field equations of a particular sort does not imply that this solution is *necessarily* physically realisable: think of the case of a pendulum with negative length. More to the point, there are as yet no solutions to the field equations that have been used to make physical predictions that are

There is, however, a potential reprieve for the A-theorist in this case. The set of spacetime solutions to Einstein’s field equations that can be foliated into spacelike hypersurfaces have taken on considerable significance over the last half a century.²⁶ The restriction to globally hyperbolic spacetime solutions is required for the Hamiltonian formulation of general relativity and this in turn is integral to using canonical quantisation techniques to develop a quantum theory of gravity. Thus, it may turn out that a successful quantum theory of gravity provides independent evidence that our spacetime is indeed globally hyperbolic, thus admitting the existence of Cauchy surfaces. This would ensure that each individual worldline in spacetime consisted of a totally ordered linear subset of the manifold, which would rekindle the possibility that each worldline is a vehicle of objective flow.

The catch, however, is that this reprieve is only plausible if it is possible to find a physical basis for fixing a preferred foliation of the spacetime manifold, which is a difficult task to say the least for a foliation invariant theory such as general relativity. A suggestion has been made in recent times, however, that the so-called *constant mean curvature* (CMC) foliation approach provides just this: a unique foliation for a reasonably large subset of spacetime solutions, which are determined by constraining the possible ways that Σ is permitted to be embedded in \mathcal{M}^4 .²⁷ This is achieved by expressing the content of the Einstein field equations in terms of the Hamiltonian constraint equations that we obtain when the canonical variables are the 3-metric and extrinsic curvature of Σ . We can then define a subset of the spacetime solutions by the condition that the mean of the extrinsic curvature is constant across Σ . As it happens, parametrising the hypersurfaces of a spacetime by constant mean curvature leads to a unique foliation of the spacetime. The A-theorists hopes for a potential reprieve, then, are pinned to whether or not a physical basis for privileging CMC foliation can be found.

To conclude this chapter I wish to briefly remark on two significant issues that foreshadow the A-theorists program in connection to the privileged foliation issue. On the bright side for the A-theorist, Belot (2007) argues that the CMC foliation approach may be an instrumental ingredient in solving the *problem of time* in general relativity. However, he also concedes that the approach “violates the spirit of general

not globally hyperbolic. Such argumentation may alleviate the worry an A-theorist might have with Gödel-type universes in the first place.

²⁶See Dirac (1958), Bergmann (1961) and Arnowitt, Deser and Misner (1962).

²⁷See Wüthrich (2010).

relativity” in that it reinstates a privileged distinction between time and space (2007, p. 219). On the not so bright side for the A-theorist, Wüthrich (2010) sets out a rather comprehensive and convincing argument against the possibility of using the CMC approach to support a particular A-theory of time: *presentism*. Thus while it seems that the A-theorist *may* find supporting physical structure in the Hamiltonian formulation of general relativity, there are significant obstacles still to be overcome.

Timelessness in Machian Gravity

We have thus far considered the picture of time that arises from pre-relativistic classical mechanics and from both the special and general theories of relativity. In this final chapter of Part I we consider the picture of time that arises from Julian Barbour's Machian formulation of general relativity and his interpretation of quantum gravity. Barbour's views are of particular interest in the current context due to his claim that they are *timeless* theories.¹

3.1 Introduction

Barbour (1994a,b, 1999) claims that both his Machian formulation of general relativity and his interpretation of canonical quantum gravity are timeless. Although Barbour's views have been scrutinised to some extent by the theoretical physics community, they seem to have been somewhat overlooked by the philosophical community.² The purpose of this chapter is to examine the extent to which the picture of reality arising from Barbour's Machian formulation of general relativity and his interpretation of canonical quantum gravity is timeless. To this end, I explore Barbour's work with a view to differentiating two senses of timelessness from the two parts of his project. I argue that we have reason to be suspect about Barbour's claim of timelessness with respect to his Machian formulation of general relativity but that he is on much firmer ground with respect to his interpretation of quantum gravity.

¹Part of this chapter consists of my contribution to the collaborative research paper Baron, Evans and Miller (2010). Any other material included here from this paper is referenced accordingly. I thank my coauthors for the opportunity to reproduce some of this work.

²Some notable exceptions include Butterfield (2002), Healey (2002), Ismael (2002), Rickles (2006) and Rickles (2008).

This chapter proceeds as follows. In §3.2 we revisit the narrative from Chapter 1 and introduce the Jacobian formulation of classical mechanics. I then motivate the two major parts of Barbour’s project. The first part consists of an argument that classical general relativity can be formulated in a Machian, and thus in some sense timeless, fashion. I present this in §3.3. The key to this argument is the claim that general relativity is an implementation of a dynamical theory that can be formulated in terms of a Jacobian geodesic principle on the space of all possible relative configurations. In the second part of his project, Barbour examines a theory of quantum gravity constructed via the quantisation of this formulation of general relativity, his interpretation of which is itself also timeless. I present this in §3.4. The key to this second part is the proposal that the Wheeler-DeWitt equation, the timeless dynamical law of canonical quantum gravity, can be interpreted as a probability distribution, defined in terms of the relative configurations, that concentrates the quantum mechanical probability on ‘time capsules’. In §3.5 I spell out the two differing senses of timelessness that we find in the two parts of Barbour’s project in terms of the formal characterisation of the features of time that was introduced in the last chapter (§2.5).

3.2 The Jacobian formulation of classical mechanics

Our point of departure in this current chapter is the point at which we arrived at the end of Chapter 1. Recall that we examined in Chapter 1 the dynamical picture of reality that arises in pre-relativistic mechanics, including both the Lagrangian and Hamiltonian pictures. We begin in this chapter with yet another formulation of classical mechanics, based on the action principle developed by German mathematician Carl Jacobi, which provides a novel perspective on the nature of time in classical physics. It is from this Jacobian theory that Barbour develops his Machian formulation of mechanics, and thus his Machian formulation of general relativity. Let us commence here by returning to Lagrangian mechanics.³

To begin with, recall that we are able to characterise Hamiltonian dynamics completely using the geometric structure of the Hamiltonian phase space, Γ ; Hamilton’s equations equip Γ with a symplectic structure through the symplectic two-form, ω . The Lagrangian configuration space, Q , likewise has a characteristic geometric

³The exposition here once again follows Lanczos (1970).

structure; the kinetic energy, T , can be used to define a line element, ds^2 , on Q which endows Q with a Riemannian structure. In curvilinear coordinates the kinetic energy becomes $T = \frac{1}{2}a_{ik}dq_i dq_k$, with the a_{ik} functions of the q_i . If we define the line element $ds^2 = 2Tdt^2 = a_{ik}dq_i dq_k$, we see that the a_{ik} play the role of a Riemannian metric tensor on Q .

With this in mind, consider once again the Lagrangian function, \mathcal{L} , of Lagrangian mechanics. For a conservative mechanical system (i.e. for a system in which the total energy is constant, $T + V = E$) whose Lagrangian is not explicitly a function of time, we can consider the time t to be a dependent variable that, along with all the q_i , is a function of some independent parameter τ . In such a case the action integral of Lagrangian mechanics (1.5) becomes (where derivatives with respect to τ are denoted by a prime):

$$A = \int_{\tau_b}^{\tau_a} \mathcal{L} \left(q_i, \frac{q'_i}{t'} \right) t' d\tau. \quad (3.1)$$

Since we find here that \mathcal{L} is a function of t' , but not t , we can employ a general reduction procedure to eliminate t as a variable and thus reduce the mechanical problem by one degree of freedom. Using the expression for the kinetic energy in terms of our new independent variable, τ , along with the relation $T = E - V$, we can derive an expression for t' :

$$t' = \frac{1}{\sqrt{2(E - V)}} \frac{ds}{d\tau}. \quad (3.2)$$

The general reduction procedure then prescribes that the reduced action integral becomes,

$$A_r = \int_{\tau_b}^{\tau_a} \sqrt{2(E - V)} ds. \quad (3.3)$$

The physically realisable dynamical trajectories through Q are then those paths for which the reduced action A_r becomes stationary; this variational principle, $\delta A_r = 0$, is called Jacobi's principle and provides another formally equivalent formulation (under appropriate conditions) of classical mechanics.

The reduced action integral A_r is not a function of the time t , but rather an arbitrary independent parameter τ , which we can simply take to be one of the q_i de-

scribing the physical system.⁴ To this extent, the dynamical picture that arises from this Jacobian formulation of classical mechanics differs from both the Lagrangian and Hamiltonian pictures. Jacobi's principle determines the motion of a physical system through the configuration space without requiring external specification of *how* this motion is described with respect to time. The dynamical behaviour of the system with respect to time can be determined by integrating (3.2) to find t as a function of τ , but this is not a part of the Jacobian variational problem, rather it is merely an expression of the conservation of energy constraint. Whereas in the Lagrangian picture we determine dynamical behaviour by completely specifying the configuration of a physical system at some definite initial and final time, in the Jacobian picture we instead completely specify the initial and final configurations of the system and then the dynamical behaviour through time is *defined* by the energy constraint. In this way, the conception of time that we find in Jacobi's formulation of classical mechanics, while maintaining the Newtonian notion of a constant monotonic time parameter, eliminates the external imposition of this parameter; time is defined internally in the configuration space.

Another significant feature of the Jacobian formulation of classical mechanics is that it enables explicit use of the Riemannian structure of configuration space to solve dynamical problems. Jacobi's principle does not merely select those paths characterised by a critical action, it determines the *shortest path* through Q between two definite end points according to the Riemannian metric; Jacobi's principle is thus a *geodesic* principle on Q .

It is this conception of time and dynamics for physical systems that is the starting point of Barbour's theory (1994a). The physical system that we will be concerned with here is the universe itself and hence we consider Q to be its corresponding configuration space. We assume that the energy of this system is both fixed and conserved, and thus we can use Jacobi's principle to determine a unique trajectory through this configuration space; the entire trajectory we take to represent the universe. Consequently, we can dispense with absolute time in the Newtonian sense and rely on the Jacobian picture to define a unique time for the universe, for which Barbour adopts the term *ephemeris time*.

⁴The ds factor of the reduced action integral does not correspond to the differential of an independent variable; it is merely the line element of Q .

3.3 Machian dynamics

If we consider the task of determining the dynamical behaviour of the universe to be an n -body problem, the configuration space of the universe, Q , has $3n$ dimensions. One can move from Q to the *relative configuration space* of the universe, Q_0 , which has $3n-6$ dimensions, by factoring out the six frame variables that specify the centre-of-mass coordinates and the orientation of the system; thereby one can remove any absolute frame from the description of the universe. The description of a system as a Jacobian theory in relative configuration space is called by Barbour *Machian*, after the Austrian physicist and philosopher Ernst Mach. By removing any notion of absolute frame from Q_0 , we are no longer able to utilise the frame dependent notion of energy to define the line element ds^2 . Thus constructing a Jacobian theory of dynamics on the relative configuration space requires that we formulate a new definition of distance between points in Q_0 . Such a definition of distance would give us, of course, a metric on Q_0 and we could then exploit the Riemannian structure of the space to determine the dynamical behaviour of the system. Barbour achieves this in the following way (1994a, p. 2862).

Imagine two configurations $\phi_1, \phi_2 \in Q_0$, with ϕ_1 described by some coordinate system q_i and ϕ_2 described relative to ϕ_1 , $q_i + \delta q_i$; the δq_i represent some arbitrary distance measure between ϕ_1 and ϕ_2 . By varying the coordinate system used to describe the two configurations we vary the measure of distance between them. We can then construct a variational principle over this distance measure, the minimum of which represents the *intrinsic difference* between the two points ϕ_1 and ϕ_2 , quantified by a Pythagorean least-squares fit. Following Barbour, let us call this process *best matching*: we can think of this process as like placing two relative configurations on top of each other and then supposing them moved relative to each other until the intrinsic difference is least (this is the essence of the Machian formulation). The intrinsic difference can then be used to define a line element, ds_0 , on Q_0 , and thus a new action integral resembling (3.3) that leads to a variational principle just like Jacobi's principle, except now we have a Machian line element. Such a principle determines geodesic paths through Q_0 that represent physically realisable four dimensional sequences of three dimensional relative configurations of the universe.

This Machian formulation of mechanics recovers Newtonian mechanics under appropriate conditions:

absolute space and time are recovered as operational concepts from the relative configurations by ‘placing’ the configurations on top of each other in the best-matching positions (*horizontal stacking*, which gives the positions in the Newtonian [centre-of-mass] system) and ‘spacing them apart’ (*vertical stacking*) in accordance with their ephemeris time differences. Thus, time and frame are obtained from a timeless and frameless ‘heap’ of relative configurations, in which all that is concrete resides. (1994a, p. 2863)

Thus far we have only considered pre-relativistic classical mechanics where points in Q_0 correspond to instantaneous relative configurations of n bodies in Euclidean space. The extension towards relativistic mechanics is achieved by considering points in the relative configuration space to correspond to 3-geometries, which are equivalence classes of Riemannian 3-metrics related by the action of the three dimensional diffeomorphism group.⁵ The notion of best matching is extended to these 3-geometries in such a way so as to define a metric on this space. Such a metric then allows the formation of an action integral with which we can construct a Jacobian principle, the stationary points of which describe physically realisable four dimensional trajectories as sequences of 3-geometries. Following Barbour (who follows Wheeler) we call this new relative configuration space *superspace* and we call the metric defining the distance between 3-geometries in superspace the *dynamical supermetric* (as opposed to the ordinary 3-metric within each 3-geometry); the relativistic Machian theory of mechanics we call *Machian geometrodynamics*. Given a sequence of 3-geometries, we can construct a four dimensional space where the instructions for layering the three dimensional hypersurfaces (3-geometries) are contained in the dynamical supermetric. We can introduce a local analogue of ephemeris time at each point on a particular hypersurface that can be used to determine the ‘temporal distance’ between that layer and the next (at the particular point in question).

Barbour claims that general relativity is just such a relativistic Machian theory of mechanics.⁶ Since general relativity is in fact foliation invariant (recall §2.6), we find that Machian geometrodynamics has a further remarkable property. If we take a sequence of 3-geometries from a geodesic in superspace and construct a four

⁵The three dimensional diffeomorphism group is the group of all bijective maps (diffeomorphisms) of a 3-metric to itself. Those 3-metrics related by diffeomorphisms are elements of the same equivalence class.

⁶Barbour’s Machian formulation of general relativity is not without its technical problems; see Pooley (2001).

dimensional metric space as above, we implicitly specify a distinct foliation of this four dimensional space; namely, the foliation determined by the 3-geometries used to construct it. If we then arbitrarily foliate this four dimensional space into a new sequence of 3-geometries we find that this new sequence is also a geodesic through superspace. This means that any

two 3-geometries in superspace will not be connected by a unique geodesic but rather a whole *sheaf* of geodesics corresponding to all possible different ways of foliating a given spacetime. (1994a, p. 2869)

Thus combining a sequence of 3-geometries from superspace amounts to constructing a four dimensional ‘spacetime’. If this constructed spacetime is in fact Lorentzian, and the 3-geometries of the sequence are all spacelike hypersurfaces, then the generalised ephemeris time introduced to determine temporal distance turns out simply to be local proper time.

Barbour claims that general relativity is not only timeless, in the sense that it can be construed as a Jacobian geodesic principle on a configuration space, it is also frameless, in the Machian sense that the configurations alone contain all the requisite information to determine any ‘temporal evolution’ between them. Time obviously is not present in any individual three dimensional relative configuration and only through the Machian principle on the relative configuration space can any semblance of time as an ordering of the instants along a geodesic be reconstructed. Due to this formulation, Barbour contends that the fundamental property of general relativity is that it is a frameless and timeless theory of the relationships of 3-geometries and superspace is the arena in which we should fundamentally describe reality. We will consider this claim of timelessness in more depth below. Firstly, though, let us turn to the second part of Barbour’s project, his timeless interpretation of canonical quantum gravity.

3.4 Canonical quantum gravity

General relativity constructed as a classical Hamiltonian theory on state space is called *canonical gravity*.⁷ A general method for quantising a classical Hamiltonian

⁷The pioneering work of casting general relativity into canonical form was carried out by Bergmann (1949, 1961), Pirani and Schild (1950), Dirac (1958), Peres (1962), DeWitt (1962) and Arnowitt, Deser and Misner (1962)

theory, a technique which has come to be known as *canonical quantisation*, was set out by Dirac (1964) in a series of lectures delivered in Canada in the 1950s and later published in his *Lectures on Quantum Mechanics*. A detailed account of this process would be too large a digression at this point but, following Pullin (2003), I provide here a sketch of Dirac's method.

Recall once again Hamilton's equations (1.9). In deriving these equations (§1.4) we assume that the classical variables, the q_i and p_i , are independent of one another. In contrast, when there exist certain relations between these variables, which we call *constraints*, the equations of motion that we derive are similar in form to Hamilton's equations but contain extra terms. It so happens that, since the classical variables are known to form a *Poisson algebra*, we can incorporate these extra terms neatly if we rewrite the constrained Hamiltonian equations of motion using the *Poisson bracket* relations. We then *quantise* this theory by making our dynamical variables operators acting on a space of wavefunctions and satisfying commutation relations corresponding to the Poisson bracket relations, and then promote the constraints to operators, which serves to restrict the class of possible wavefunctions.

Barbour's Jacobian theory can be transformed into a Hamiltonian theory by defining the canonical momenta in the standard way (1.6). When we do so, we find that there indeed exists a constraint between the q_i and p_i and that the resulting equations of motion can be written in terms of the Poisson bracket relations.⁸ We can then follow Dirac's constrained Hamiltonian quantisation procedure to quantise the theory. It so happens that when we subject the Hamiltonian formulation of the *pre-relativistic* Jacobian theory to this procedure the dynamical equation that results is the time-independent Schrödinger equation of nonrelativistic quantum mechanics (we will meet the orthodox formulation of quantum mechanics in Chapter 4). Applied to Machian geometrodynamics, this recipe produces the Wheeler-DeWitt equation⁹: a constraint equation of the form $\hat{H}\Psi = 0$, where Ψ is a complex-valued functional of 3-geometries (i.e. the points of superspace) (Butterfield and Isham, 1999, p. 149). When the canonical formulation of general relativity is subject to such canonical quantisation techniques, the theory of quantum gravity obtained is referred to as *canonical quantum gravity*.

The goal of this section then is to introduce the picture of reality that Barbour

⁸Gryb (2010) contains a clear and concise presentation of this process.

⁹This was first carried out by Peres (1962).

envisages as arising from his interpretation of canonical quantum gravity (1994b). As a preamble to his interpretation Barbour emphasises that both general relativity (in his Machian formulation) and quantum theory can be characterised as a theory on relative configuration space: the Schrödinger wavefunction of any system is defined over all its possible configurations, and thus instead of describing a unique classical history in the configuration space of the system, the quantum wavefunction explores all configurations. This leads Barbour to suggest that the Wheeler-DeWitt equation can be interpreted as describing a static wavefunction Ψ on superspace.

On this view, the notion of a Hilbert space representing the state space of some subsystem of the universe is simply redundant; Barbour proposes to treat the universe as a single holistic quantum system. In any one configuration, no distinction is possible between quantum system and measurement device: all are simply part of a particular configuration of the universe. The sole role of the wavefunction, as in Born's probability interpretation (§4.3), is to say how likely the actualising of a given configuration is. These probabilities are not, however, time dependent nor are they conditioned on prior knowledge and tied to measurement setups; they are given once and for all for the possible configurations that the universe could be in. It is in this sense that Barbour's interpretation of canonical quantum gravity is timeless.

In trying to motivate an intuitive picture of his model, Barbour contrasts two ways that we might imagine such a universe (1994b, p. 2881). On the one hand, from an external viewpoint, we can imagine the set of all relative configurations of superspace to exist as a *heap of possibilities*.¹⁰ We can then divide superspace into infinitesimal hypercubes, take the value of Ψ in each hypercube, calculate $\Psi\Psi^*$ and place a number proportional to $\Psi\Psi^*$ of identical copies of a representative configuration of that hypercube into a second heap called the *heap of actualities*.¹¹ We may now suppose that drawing one configuration at random from the heap of actualities actualises that configuration. Thus, a probable configuration is more likely to be actualised than one that is improbable.

On the other hand, from an internal viewpoint, we have it that our direct experience, including that of motion, is correlated only with configurations of our brains.

¹⁰Barbour emphasises that this is called a 'heap' because each point in the relative configuration space has, unlike an ordinary manifold, an individual existence outside the space, i.e. a three dimensional configuration.

¹¹In what way we are to imagine this heap of actualities is unclear. I abstain from exploring this issue.

“Our seeing motion at some instant is correlated with a *single* configuration of our brain that contains, so to speak, *several* stills of a movie that we are aware of at once and interpret as motion” (1994b, p. 2883). The connection between the internal and external views is that while some “divine mathematician” actualises (by random selection) one particular configuration of the universe, it seems to us as though we are inside part of that configuration and have direct awareness of that part as an experienced instant. The problem in orthodox quantum theory concerning the reality of the unactualised possibilities is compounded in Barbour’s quantum gravity

since one even has to ask whether events of which we have vivid memories are actually experienced. This is because everything we experience in any instant, including the memories themselves, must be coded in our instantaneous brain configuration. Records of apparent past events are in fact details in the present configuration. And all the timeless theory tells us is that each such configuration has a certain probability. (1994b, p. 2883)

Thus while we have direct evidence that the present configuration is actualised, we are epistemically locked in this configuration and therefore have no warrant for believing that any other instant is actually experienced. Barbour’s quantum gravity “does seem to come perilously close to solipsism of the instant” (1994b, p. 2883).

The most significant element of Barbour’s interpretation of quantum gravity is the notion of a time capsule. A time capsule is a static configuration of part or all of the universe containing structures which suggest they are mutually consistent records of processes that took place in a past in accordance with certain laws. Time capsules as a bare possibility exist, and it is the existence of such special configurations that Barbour claims allow us to recover the appearance of time from a timeless picture of reality. However, the set of all time capsules has negligible measure amongst the set of all possible configurations, thus Barbour’s proposal is conditional upon his suggestion that the solution to the Wheeler-DeWitt equation concentrates the quantum probability distribution on time capsules, thereby making it probable that we would find ourselves in a three dimensional configuration that contained evidence of having been created by a dynamical process.¹²

¹²Barbour argues that this suggestion can be supported by extending the Mott (1929) formalism for α -particle scattering. While this is certainly integral the cogency of Barbour’s view, I leave this issue to one side in the present discussion and take for granted the claim that the Wheeler-DeWitt equation concentrates the quantum probability distribution on time capsules.

One advantage of Barbour's interpretation of canonical quantum gravity comes in the form of a possible explanation for two notoriously unexplained (and related) problems of cosmology: namely, the directionality of time and the improbable initial conditions of the universe. Both might be explained if it turns out that the solution to the Wheeler-DeWitt equation picks out as probable configurations on superspace that carry with them records of having evolved from a low entropy past. Interestingly, Barbour suggests that the topological structure of superspace might play a significant role in any such explanation. If we consider a three dimensional relative configuration space describing the relative configurations of a system of three point particles free to move in three spatial dimensions, we find that actual configurations of this system can only be found in a portion of the positive octant of the configuration space. By analogy with this case Barbour surmises that superspace must have a similarly asymmetric structure. Therefore, since spacetime and initial conditions are irrelevant in superspace, and only a static balance between relative configurations can be called upon to account for the arrow of time, the question of time's directionality and the improbable initial conditions of the universe boil down to the question of what sort of configurations attract large values of the quantum mechanical probability. The structure of superspace then "cannot but give rise to an intricate and inherently asymmetric topography" (Barbour, 1994b, p. 2893) for the potential of the quantum wave equation.

Furthermore, it is apparent that there is a natural origin of superspace which represents the configuration in which everything sits upon everything else. It is not lost on Barbour that our current standard cosmological models suggest that the universe has emerged from a tightly compressed state of very high density at a finite time in the past. A timeless interpretation of this fact in terms of superspace is that the Big Bang is not in the past but it is merely a special configuration in superspace. Furthermore, due to the characteristic structure of superspace, any particular configuration is a finite distance from an absolute frontier to one side (the origin) while there is no frontier to the other side.

For, at least in principle, the absolute frontier... could have the effect of concentrating the wavefunction of the universe on configurations that seem to contain records... of configurations that lie between the recording configurations and the absolute frontier but none of configurations that lie 'beyond' the recording configurations. (1994b, p. 2893)

Barbour sums up the moral of the second part of his project in the following way: “Time is not a framework in which the configurations of the world evolve. Time exists only so far as concrete configurations express it in their structure” (1994b, p. 2885). Let us now turn our attention to Barbour’s claim that both his Machian formulation of general relativity and his interpretation of canonical quantum gravity are timeless.

3.5 From timeless physical theory to timelessness

We have thus far familiarised ourselves with the details of the two parts of Barbour’s project: both his Machian formulation of general relativity and his interpretation of quantum gravity. We see that Barbour’s claim of timelessness (that the scientific image derived from his theories is timeless) is motivated by the fact that, according to his theories, superspace is the fundamental arena with which we describe the universe, and superspace is comprised of three dimensional relative configurations. I refrain here from challenging the technical details of Barbour’s theories and pose a challenge of a different sort: given the picture of reality that arises from Barbour’s theories, to what extent is Barbour warranted in drawing his conclusion that this picture is timeless? The problematic issue in this respect, as Baron *et al.* (2010) make clear, is that Barbour is missing a crucial piece of the puzzle in his argument from ‘timeless’ physical theories to timelessness; Barbour fails to provide a characterisation of the essential features of time. Without this characterisation, it is not clear exactly which features of time must be absent from a physical theory to render the picture of reality arising from it ‘timeless’. The goal of this section then is to explore Barbour’s claim of timelessness in light of a potential characterisation of the essential features of time.

Before we embark on this exploration I introduce here some terminology that will help us distinguish the different senses of timelessness in Barbour’s work. The core of Barbour’s claim of timelessness is that the fundamental elements of our description of reality, both in his classical theory and his quantum theory, are *three* dimensional relative configurations, i.e. ‘frozen’ configurations which lack a temporal dimension. However, there is a significant difference between the timelessness of Barbour’s Machian formulation of general relativity and the timelessness of his interpretation of canonical quantum gravity. In the former sense, the timelessness of

the relative configurations is supplemented by a Machian reconstruction of temporal structure; using only the data present within the set of relative configurations we can determine geodesics through Q_0 which correspond to physically realisable trajectories, and we can define a temporal metric from these trajectories. As Butterfield (2002, p. 15) remarks, the Machian formulation of general relativity “will deserve to be called ‘timeless’, in that there is no time metric in Q_0 ; rather... the time metric is definable from the dynamics”. Thus while the theory is timeless in the sense that a time dimension is absent from the fundamental elements of the theory, a temporal metric of sorts can be reconstructed from these timeless elements. Let us call this *Machian timelessness*.

The same cannot be said of the latter sense of timelessness. The timelessness of Barbour’s interpretation of canonical quantum gravity is again manifested in the absence of a time dimension in the fundamental elements of the theory but, in contrast to Machian timelessness, this is then compounded by the remaining structure of the theory: there exists a time independent (static) quantum probability distribution (QPD) across the relative configuration space that is concentrated upon time capsules, i.e. special three dimensional configurations that merely *appear* as though they have been created from a dynamical process. Thus in quantising the Machian formulation of general relativity to yield Barbour’s particular interpretation of quantum gravity we lose an element of Machian temporal reconstruction and gain an account of temporal appearances in the form of time capsules. Let us call this sense of timelessness *QPD timelessness*. Barbour does not provide a clear statement distinguishing these two senses of timelessness. Indeed, Butterfield (2002, p. 3) again, “the book [The End of Time] gives the misleading impression that Barbour’s various views are closely connected one with another”. The distinction between these two senses of timelessness will become more clear through our consideration of the essential features of time.

The characterisation of the essential features of time that we consider here is one that we have encountered previously (in both the Introduction and in §2.5): Rovelli’s analysis of the different notions of time in physical theory. Recall that Rovelli (1995, 2004) identifies up to nine distinct attributes that we assign to the temporal structure of our various contemporary physical theories and folk concepts, including directionality, uniqueness and globality, amongst others. Rovelli proposes that our contemporary physical theories can be arranged in a hierarchical structure

in which an increase in the universality of the theory corresponds to a decrease in the possible attributes that we can assign to the temporal structure of the theory. In §2.6 we adapted Rovelli's characterisation and formalised these features with respect to relativity theory and concluded that the structure of (proper) time in orthodox general relativity (one of our most universal physical theories) can be represented by the set $\{\mathcal{M}^1, <, g_{\mu\nu}\}$; that is, we find time in general relativity to be represented by a (local) ordered, linear and metric substructure of the four dimensional spacetime manifold. In this respect, according to Rovelli, the two possible attributes that we find to characterise the temporal structure of general relativity are *linearity* ('time' can be used to refer to a one dimensional substructure of ordered temporal instants) and *metricity* ('time' can be used to refer to the meaningful measure of distance between any two time instants).

Using this as a characterisation of the essential features of time in general relativity, we can attempt to interpret Barbour's claim of timelessness as a denial that there is some linear and metric substructure within his theories that can be referred to as time. This provides then a straightforward formula for evaluating Barbour's claim that both his Machian formulation of general relativity and his interpretation of quantum gravity entail a timeless picture of reality: for both of Barbour's theories we consider the extent to which linearity and metricity can be extrapolated from the picture of reality that arises.

In Barbour's Machian formulation of general relativity and his interpretation of canonical quantum gravity the fundamental elements of the theory are three dimensional relative configurations. If we consider a single relative configuration in isolation, there is no one dimensional substructure therein to identify as time and no way to meaningfully measure the temporal distance from this configuration to any other. Thus in both Barbour's classical and quantum theories, when we consider a single instant in the relative configuration space, we notice (quite trivially) that there is no fundamental linear or metric structure that we might identify as time.

However, an integral part of Barbour's Machian formulation of general relativity is the specific and detailed Machian algorithm that enables one to define an intrinsic measure of distance between any two points in superspace; this measure is just the dynamical supermetric. The supermetric, through Jacobi's principle, establishes the geodesics of superspace, which yield sequences of 3-geometries that correspond to the four dimensional 'spacetime' solutions of Einstein's field equations. As mentioned

before, if the constructed spacetime turns out to be Lorentzian and the 3-geometries are all spacelike hypersurfaces of this spacetime, then the ephemeris time of the Jacobian geodesics is simply the proper time of orthodox general relativity, and thus has the structure $\{\mathcal{M}^1, <, g_{\mu\nu}\}$. We see then that a linear and metric substructure that we refer to as time emerges explicitly from the Machian formulation of general relativity.¹³ Thus when it comes to Barbour's Machian formulation of general relativity, linearity and metricity are not entirely absent from the theory. The relevant features exist, it is just that they emerge out of the three dimensional relative configurations of superspace via the Machian best-matching algorithm. Thus it becomes apparent that talk of time is not misplaced in Barbour's Machian general relativity; there is temporal structure, it is just not found in the fundamental components of the theory.

Given that the superspace of Barbour's formulation of general relativity explicitly yields a temporal parameter that corresponds directly with that of orthodox general relativity, we have reason to be suspicious of Barbour's claim, with respect to his Machian formulation of general relativity, that the corresponding picture of reality that arises is timeless. There seems to be an uneasy tension between Barbour's claim of timelessness motivated by the absence of a temporal dimension in the fundamental elements of configuration space and his own statement that his theory contains explicit structure which he is happy to employ as a definition of time (see fn. 13). It appears at the very least that Barbour's claim that his Machian formulation of general relativity is timeless is slightly overstated.¹⁴

As a brief aside, a parallel argument can be constructed from more orthodox metaphysical considerations in support of this conclusion that a timeless picture of reality does not follow from the Machian timelessness of Barbour's general relativity (Baron *et al.*, 2010). Consider once again McTaggart's distinction between the A- and B-series and the subsequent temporal models that embody this distinction: the A- and B-theories of time (§2.1.1). These temporal models come prepackaged with

¹³Without these two conditions, ephemeris time does not correspond to proper time but to a measure of the four dimensional interval orthogonal to the hypersurfaces of the foliation given by the sequence of 3-geometries in question. Regardless of any correspondence to proper time, however, we still find a linear and metric substructure to arise. Moreover, we find Barbour himself stating: "This is what I propose to call time" (1994b, p. 2877).

¹⁴Of course, one option available to Barbour is to claim that being one of the fundamental posits in our best physical theory is essential to the nature of time. See Baron *et al.* (2010) for a discussion of this point.

a supposition about the essential features of time; in short, the A-theory supposes the A-series to be essential and the B-theory supposes the B-series to be essential. We might then characterise Barbour's claim of timelessness as a claim that the picture of reality arising from his theories is without one or the other of these essential features of time. If we put to one side the uninteresting case of reading Barbour as denying the reality of the A-series¹⁵, we might characterise Barbour as claiming the following: it is essential to time that there is a B-series, and since there is no such series, then there is no time. Denying the existence of the B-series, recall, is denying the existence of the temporal relations 'earlier than', 'later than' and 'simultaneous with'.

If we now consider the Machian algorithm for reconstructing four dimensional 'spacetime' solutions from the three dimensional relative configurations of superspace, we see quite clearly that this algorithm yields an explicit reconstruction of a temporal ordering via the foliation provided by the sequence of 3-geometries. This temporal ordering, of course, is simply a B-series of temporal instants. Thus if we read Barbour's claim of timelessness as a denial of the existence of the B-series, we see that Machian timelessness does not entail a timeless picture of reality. There is a close correspondence between this argument and the argument above; one might argue that they are merely terminological variants of each other. In both analyses we find that in Barbour's Machian formulation of general relativity certain essential features of time remain as integral parts of the theory: a linear and metric temporal structure, on the one hand, and a B-series, on the other. Put like this, however, it should now be clear why there is a close correspondence between these two arguments: the temporal structure that the B-series provides just is a linear and metric temporal structure, and *vice versa*.¹⁶

Consider now QPD timelessness and Barbour's interpretation of canonical quantum gravity: the task of extrapolating linearity and metricity from the picture of reality that arises here is much more difficult. The fundamental constituents of the theory are still three dimensional relative configurations in superspace. However, whereas Barbour used the geometry of superspace to recover orthodox general rela-

¹⁵Since in most cases even the A-theorist would agree that even if there is no A-series then there is still a B-series and this is enough for there to be time.

¹⁶Likewise, the temporal structure that the A-series provides might also be characterised in terms of Rovelli's attributes: linearity, metricity, globality, externality, uniqueness, directionality and presentness.

tivity (and thus ‘time’, in a sense) in his Machian formulation of general relativity, the explanation provided for our experience of time in his interpretation of canonical quantum gravity is the ‘time capsule’; the appearance of the present configuration having evolved in time is merely an illusion brought about by the mutually consistent records we find in each time capsule. Thus there is no need to specify some algorithm for defining a meaningful measure of distance between the relative configurations of superspace, and there is no need to construct a linear ordering from the “heap” of three dimensional instants. Not only is it the case then that there is no linear or metric structure in the theory, such structure is not recoverable from superspace in any sense. At best, there is the mere illusion of linear and metric structure via the time capsules.¹⁷

We see then that the QPD timelessness at the heart of Barbour’s interpretation of quantum gravity does entail that certain features of time that we might take to be essential (namely, linearity and metricity) are indeed absent from the theory. Barbour is then on much firmer ground when he claims that the picture of reality that arises from his interpretation of canonical quantum gravity is timeless.¹⁸ Consequently, this exacerbates the tension that is problematic with respect to Machian timelessness: without the more solid claim concerning QPD timelessness, Barbour may have been able to argue that Machian timelessness (as I have characterised it here) is simply what he meant all along by timelessness. In the present case, however, this would render his claim of QPD timelessness rather *understated*, since we have much more reason to believe that the picture of reality that arises in this context really is timeless.

Thus while we have reason to be suspect of Barbour’s justification that the picture of reality arising from his Machian formulation of general relativity is timeless, he is much more justified in claiming a timeless picture of reality to arise from his interpretation of canonical quantum gravity.

¹⁷For an interesting discussion of structure that might be associated with the set of time capsules see Healey (2002).

¹⁸In terms of the aforementioned terminological variant we can characterise the absence of any linear and metric structure between points of superspace as a set of time instants that are not related by the temporal relations ‘earlier than’, ‘later than’ or ‘simultaneous with’. Baron *et al.* (2010) liken Barbour’s position here to McTaggart’s position that all that exists is a C-series of time instants that contains no temporal ordering whatsoever. Both Barbour and McTaggart conclude that there is no time.

★ ★ ★

This draws Part I to a close. The intention here has been to demonstrate what an analysis of the nature of time amounts to when we take seriously the doctrine that modern physics should be treated as the primary guide to the nature of time. The conclusions that we have reached are unsurprisingly counterintuitive; modern spacetime physics no longer has the intuitive appeal of the Newtonian picture. In fact, even within the scope of analytical mechanics we find a novel and interesting picture of reality arising from a Lagrangian picture of teleological determination. We have also seen that relativity theory provides constraints on the metaphysical possibilities for time and we have become familiar with the nature of time within a Machian formulation of general relativity as well as an interpretation of canonical quantum gravity. We shift our attention now away from spacetime physics and toward another of the pillars of modern physical theory: quantum mechanics.

Part II

Quantum Theory and the Newtonian Picture of Time

Quantum Mechanics, EPR and Escaping Bell's Theorem

We have thus far examined the picture of time that arises in various physical theories and we have considered the way in which this picture impinges upon metaphysical considerations concerning the nature of time. One of the morals of Part I was that metaphysical considerations concerning the nature of time must respect the constraints and limitations imposed by the picture of time that arises in physics; we should treat physical theory as an authority and primary guide in this respect. In Part II we turn our attention to a different sort of physical theory and a different sort of metaphysical consideration and address a confusion that can be seen as arising in the absence of the methodology in Part I. Specifically, study into the nature of time should be *guided* by modern physics and thus we should be careful not to *insert* any preconceived Newtonian conception of time unwittingly into our interpretation of the formalism of physical theory.

Our central concern for the remainder of this thesis will be the interpretation of the mathematical formalism of nonrelativistic quantum mechanics. We can think of the task of providing an interpretation of quantum mechanics as commensurate with the task of investigating the picture of reality that arises from the quantum formalism. However, we face a problem building such a picture of reality from the quantum formalism: the theory of quantum mechanics is in certain crucial respects different from any theory of classical mechanics that precedes it and thus much of the Newtonian picture of reality must give way to a new picture. The task of deciding exactly which of our Newtonian intuitions, if any, is worth keeping is notoriously difficult since it is the case that a complete quantum picture of reality is underdetermined by the formalism. Moreover, some of these Newtonian intuitions

are so engrained in the way we think about the world that often a quantum picture of reality is unwittingly built atop a whole raft of preconceived Newtonian metaphysics.

The goal of this second part of the thesis then is to explore the interpretation of nonrelativistic quantum mechanics in light of the contention that a Newtonian conception of time besets the orthodox quantum picture. In particular, I argue in favour of introducing backwards-in-time causal influences as part of an alternative conception of time that is consistent with the quantum picture of reality. The argument proceeds as follows. In this chapter I introduce the formal structure of quantum mechanics and the accompanying interpretational debate. I aim to show that retrocausality (backwards-in-time causality) can be instrumental in alleviating the interpretational difficulties posed by Bell's theorem. I provide in the next chapter an analysis of the metaphysical possibilities that our current scientific theories allow to mount an independent argument in favour of retrocausality; namely, that retrocausality cannot be ruled out as a metaphysical position on analytic grounds. In Chapter 6 I consider Maudlin's argument against retrocausality, motivated by his critique of the transactional interpretation of quantum mechanics, which I claim is a clear exemplar of the misapplication of a preconceived Newtonian conception of time. Let us begin, however, with the theory of quantum mechanics.

4.1 Introduction

One of the great achievements of 20th century physics is the development of the theory of quantum mechanics. One could argue, however, that it is a minor embarrassment that there is no universally accepted interpretation of quantum theory that provides a coherent quantum picture of reality, despite the many attempts to do so. As mentioned above, any attempt to provide this picture must abandon at least some of the Newtonian picture of reality. The goal of this chapter is to motivate the introduction of retrocausality as one such attempt to provide this coherent picture. This chapter follows a somewhat historical narrative, which serves two purposes: the first is to explain how the proposal of retrocausality fits in to the debate about the interpretation of quantum mechanics; and the second is to do this in a way which portrays retrocausality as the 'lesser-of-two-evils' resolution to the interpretational difficulties that arise in this debate.

I begin in §4.2 by introducing the core principles of the *mechanistic worldview*

of classical mechanics, which shares much in common with the Newtonian picture of reality. I introduce in §4.3 the fundamental formalism of quantum mechanics alongside the interpretational problems that this creates for the mechanistic worldview. I then introduce the Copenhagen interpretation in §4.4, which can be seen as the first interpretation of quantum mechanics to become an orthodoxy. Following this in §4.5 I present the details of an argument set out by Einstein, Podolsky and Rosen (EPR) against the Copenhagen interpretation. This analysis leads us to Bell's theorem in §4.6 and it is at this point that the hypothesis of retrocausality can be employed as a way of escaping Bell's result. I finish the chapter in §4.7 by providing a sketch of a retrocausal description of the strange correlations we find in the sorts of quantum systems at the heart of Bell's analysis. Our narrative must begin somewhere, however, so let us begin with classical physics at the turn of the 20th century.

4.2 Classical physics

Physics prior to the 20th century was driven by a mechanistic worldview; recall the picture of reality that we found to arise from classical mechanics (Chapter 1). This view is borne out of the development of Newtonian physics in the 17th century and grew throughout the reformulation of classical mechanics in the 18th and 19th centuries as well as the theory of relativity at the beginning of the 20th century. The success of analytical mathematics in describing physical systems was instrumental in engendering a belief in the scientific community of this era that there are strict natural laws that govern the universe as a whole and that these laws can be formalised into mathematical systems. The mathematical formalism describing physical systems thus motivates a rigorous picture of reality according to classical mechanics. Among the various elements that characterise the mechanistic worldview, two in particular warrant our close attention and will be continually within our sights for much of Part II: the generative picture of determination and the principle of locality. Let us begin with the generative picture of determination.

Recall that in classical mechanics the dynamical behaviour of a physical system is described by differential equations ((1.4), for instance). One feature of this Newtonian schema is that the complete specification of the state of a physical system at some time determines, through the differential equations governing the dynamics

of the system, all past and future states of the system. If one then accepts the mechanistic doctrine that the universe as a whole can in principle be described in terms of formal mathematical systems, then this feature of physical systems should extend to the universe as a whole. That is, if one could specify the exact state of the universe at any one time and the dynamical laws governing the universe, then one could establish the state of the universe at all past and future times. This is the *generative* picture of determination (recall §1.2).¹ This picture captures the image of the universe as a clockwork machine which invariably and eternally operates as a function of the inner cogs and gears of the machine. Within such a picture of reality the only limitation set on predicting the future with certainty would be our capacity to measure and then process such enormous data sets such as the exact state of the universe at any moment in time.

We should note here that, depending on our particular views about causality, this picture of determination might render our causal concept redundant. If the state of the universe at any one time were enough to determine the state of the universe for all times given the dynamical laws of the universe then any mechanistic account of causality consistent with this would merely be a restatement of this very fact; the state of the universe at any time would be by definition *caused* by its state at some given time slice. This identification of mechanistic causation and determinism can be seen as one of the central pillars of the mechanistic worldview. (In the next chapter, §5.2.2, I introduce an account of causality that strikes a balance between a deterministic view of the universe and our intuitive notion of causality. We will return to this issue at various times throughout Part II.) Let us now consider the principle of locality.

Since the principle of locality is to play a crucial role in the motivation and analysis of retrocausality, it will serve us well at this point to spend some time getting clear on what we mean by 'local'. We will mostly be concerned here with the principle of locality as it is understood with respect to the theory of relativity. However, to get a feel for what it is in the world that the principle of locality is attempting to capture I wish to begin with an account of locality as it might be characterised from within the mechanistic worldview.

Recall that in classical mechanics we represent the dynamical behaviour of a

¹We could also refer to this as the Laplacian picture of determination, after the French mathematician Pierre Simon Laplace (Earman, 1986).

physical system in a state space, where each point of the space represents some possible combination of values for the generalised coordinates we use to describe the system. Recall also that a history of a physical system is represented by a path through this state space determined by the dynamical laws of the system, with each point along the curve representing the state of the system at some particular moment in time through its history. Due to the nature of the differential equations which describe these paths through state space, the values of the generalised coordinates describing the system evolve along the path (through time) in a continuous manner; it is simply not possible according to the mathematical formalism that there be discontinuous jumps in the values of the generalised coordinates characterising a physical system from one time to the next. In other words, the mathematical formalism of classical mechanics forbids matter and energy to be spontaneously destroyed at one spacetime location only to be created again at another distant spacetime location.

In very general terms, the continuity of the dynamical behaviour of a physical system according to classical mechanics can be seen as capturing the everyday intuition that the behaviours of the physical systems we observe are *causally* connected in a particular way. It is not the case that we experience the sorts of uncaused behaviours that discontinuous state space jumps would imply of the classical world (such as matter appearing from nothing or moving objects suddenly stopping without any external forces). In a rudimentary sense, then, one is left with the impression that the causal structure of the four dimensional spatiotemporal manifold we inhabit (represented by a path through state space) is *local* in the sense that the behaviour of some physical system is influenced only by the events in its immediate spatial and temporal vicinity. While we are thus far only considering theories of classical mechanics, this is the sort of intuition about the nature of reality that we would like to capture with the principle of locality - a kind of *action-by-contact*.²

The principle of locality was formalised and extended with the introduction of the special theory of relativity, which is a theory about the connectedness of space and

²Of course, we find at least one notable, perfectly good classical theory that is itself nonlocal in this action-by-contact sense: Newtonian gravity. Although there was in fact never a successful theory of gravity as part of the mechanistic worldview that failed to violate this notion of action-by-contact, Newton's own displeasure with the action-at-a-distance of his theory of gravity suggests this notion is not unimportant to the conceptual foundation of classical physics.

time themselves (recall §2.2).³ Einstein's characterisation of his theory states that it is light signals that provide this connection between separated parts of spacetime and this accordingly places an upper bound on the spatiotemporal distance between causally connected, or local, events. Thus given a particular spacetime location, there is a precise account according to relativity theory of which other spacetime locations are causally connected (and are therefore local) and which spacetime locations are not causally connected (and are therefore nonlocal).⁴ Recall from Chapter 2 that the former class of spacetime locations are said to be timelike separated from the given location and the latter are said to be spacelike separated. The principle of locality, then, is in essence a principle concerning the causal connectedness of the spacetime manifold by which physical systems are constrained.

The generative picture of determination and the principle of locality grew from the conceptual framework of the mechanistic worldview. However, physics in the 20th century was fundamentally to challenge this worldview. The continuing development of thermodynamics and electromagnetism towards the end of the 19th century led to the construction of the theory of quantum mechanics. Throughout the development of quantum theory, it became apparent that these elements of the Newtonian picture do not seem to provide a good picture of the behaviour of quantum systems. Quantum mechanics had thus laid a very real challenge to the picture of reality that arises from classical mechanics. Before we can address the consequences of this challenge for the ensuing quantum picture, we must address in much more detail the challenge itself. Let us turn our attention to the foundations of quantum mechanics.

4.3 The theory of quantum mechanics

The theory of quantum mechanics is an elegant mathematical model which is used to describe physical processes that occur on an atomic scale. At its inception it was also a unique case amongst scientific theories. The reason for this is that the mathematical formalism of the theory is rather abstract and thus does not lend

³More specifically, special relativity is a theory about the connectedness of the measurements of space and time carried out by inertial observers using rulers and clocks.

⁴The local/nonlocal distinction should not be confused with local/global distinction. The former concerns timelike as opposed to spacelike separation and the latter concerns the immediate neighbourhood of a point in a space as opposed to the entire space.

itself to the sort of natural interpretation that is familiar in classical mechanics. This is not to say that the abstract mathematical formalism is somehow detached from physical reality. On the contrary, the need for a new theory of mechanics at the atomic scale over and above classical mechanics was due to the inability of classical mechanics to account for a multitude of atomic phenomena. The theory of quantum mechanics slowly replaced classical mechanics through a series of steps in response to experimental developments in the early part of the 20th century. In fact, it took roughly a generation from the time the first quantum phenomena were observed, around the turn of the 20th century, to the time an encompassing empirically adequate theory was developed in the 1920s.

At first, two such theories were developed: Schrödinger's (1926) wave mechanics and Heisenberg's (1925) matrix mechanics. Each appeared to account for the observed phenomena equally well but each emphasised different aspects of the phenomena as being the integral element of the description. This discrepancy underpins the interpretational debate in quantum mechanics. Soon after the development of these competing theories of quantum mechanics, Dirac (1930) and von Neumann (1932) were able to show that the two formulations were mathematically equivalent and by the early 1930s the theory of quantum mechanics had grown into a sophisticated piece of mathematical machinery. Unfortunately this did not provide a definitive solution to the interpretational discrepancy; more on this shortly.⁵

Before we can begin to come to grips with these interpretational difficulties, it will be useful to familiarise ourselves with the basic mathematical machinery which forms the core of the theory of quantum mechanics.⁶ I will by no means give a mathematically complete account of the formalism but I do hope to provide here enough structure on which we can start to hang a physical interpretation. Moreover, as I mentioned above, this mathematical formalism is quite abstract. While it can

⁵Of course, this was not the end of the development of quantum mechanics, either. In making quantum mechanics consistent with the special theory of relativity, the theory of quantum electrodynamics was created, which until recently was the most successful physical theory yet developed (along with the standard model of particle physics, of which it is a part; success here is measured in terms of empirical adequacy). I will not focus on these later incarnations of the theory of quantum mechanics here; while I will at certain points make reference to theories such as quantum electrodynamics and the like, the focus here will be on the quantum picture of time that arises from considerations of nonlocality in nonrelativistic quantum mechanics.

⁶To get a flavour of the interpretation of quantum mechanics in a historical context, the exposition here mostly follows Bohm (1952b).

indeed be instructive to present the formalism with one eye on the experimental developments which guided its creation, for brevity I will not pursue such a heuristic presentation here.

To begin with, we represent the state of a quantum system with a *wavefunction*, ψ , that is the solution to the Schrödinger equation (a linear wave equation describing the time evolution of a quantum system). The wavefunction is not taken to specify the exact values of all the degrees of freedom of the physical system it represents, like one would expect in classical mechanics, but rather it contains information about the probable outcome of various measurements that can be made on the system.

Since the Schrödinger equation describes the evolution of the wavefunction, and since the wavefunction contains only probabilistic information about the quantum system, we can liken the evolution of the wavefunction to the evolution of a wave packet in classical mechanics. In an evolving classical wave packet there is a spread of values for the frequencies and velocities of the constituent waves of the packet. By the same token, in an evolving quantum mechanical wavefunction there is a spread in the certainty to which values for the position and momentum of the quantum system can be measured. In fact, in quantum mechanics there is a reciprocity in this spread in certainty which holds between conjugate pairs such as position and momentum. Heisenberg's indeterminacy relation quantifies this reciprocity. One is thus unable to measure simultaneously and to arbitrary accuracy determinate values for particular pairs of properties in a quantum system.

Information about the outcome of a measurement is calculated in quantum mechanics through the introduction of an *operator* for each physically observable property of the quantum system. Thus, given a wavefunction representing a quantum system and an operator representing an observable property of that system, we can calculate what the expected (i.e. most probable) outcome of an experiment would be if we were to measure the value of the corresponding property of the quantum system. Since this information is probabilistic, the actual measurements of the system would only result in this calculated outcome on average.

In contrast, it *is* possible to have a quantum system in such a state so as to yield the same predictable outcome for some measured observable property on every such measurement. The wavefunction representation of such a system is part of a special class of wavefunctions for each operator of this sort. Consider a quantum system

represented by ψ with an observable property A corresponding to an operator A such that any possible measurement of property A yields a stable and predictable value, a , for this property. Due to the special nature of the relationship between ψ , A and a , there is a simplified method for calculating this outcome: we simply operate on the wavefunction ψ with the operator A , given by $A\psi$, to yield the result $a\psi$, i.e. $A\psi = a\psi$. For any A , ψ and a obeying this relation, ψ is said to be an *eigenstate* (or *eigenfunction*) of A with *eigenvalue* a . Thus for any operator there exists a class of wavefunctions that are eigenstates of the operator with corresponding eigenvalues. We should note here that the operators have no physical significance unless they are acting on an eigenfunction in this manner; that is, we do not ascribe to a system a definite value of the observable corresponding to an operator A unless the system can be represented by an eigenstate of A .

Both the set of all wavefunctions and the set of all operators respectively have particular mathematical properties. The set of all wavefunctions form a state space called a *Hilbert space* such that a linear combination of any wavefunctions from the space will result in a wavefunction that is also in the space. The operators, which act on the elements of the vector space, form a nonabelian algebra; that is, pairs of operators in general do not *commute*, i.e. for any two operators A and B , $AB \neq BA$. A pair of physical properties with corresponding operators, A and B , are said to be *conjugate variables* if the operators satisfy the *canonical commutation relations*: $AB - BA = i\hbar\mathbb{I}$. The noncommutativity of such operators is contained in Heisenberg's indeterminacy relation: conjugate variables cannot have determinate values simultaneously.

Since we have it that any operator defines a class of wavefunctions that are its eigenstates, and that the set of all wavefunctions forms a Hilbert space, this leads us to the result that any arbitrary wavefunction can be written as a linear combination of eigenstates of some operator. Thus, given a wavefunction representation of the state of a quantum system and an operator corresponding to some observable property of that system, the wavefunction representation can be expanded in terms of the eigenstates of that operator. As we will see in more detail in the next section, the square of the coefficient of each of the terms of this linear combination can be interpreted as the probability of finding as the value of the observable property the eigenvalue associated with that eigenstate (this is known as the Born rule). For instance, if our quantum system is represented by ψ and the observable property we

are interested in corresponds to the operator A , then we can write the expansion of the wavefunction as

$$\psi = \sum_a P_a \psi_a, \quad (4.1)$$

where the ψ_a are the eigenstates of A , each P_a is a projection onto the eigenspace corresponding to the eigenvalue a of ψ_a and the square of the modulus of each projection, $|P_a|^2$, is the probability of finding the corresponding measurement outcome to be a .

For it to be the case that a measurement outcome, say value a of property A , is actualised, the state of our quantum system must be represented by an eigenstate of the corresponding operator A , say ψ_a . Thus according to the formalism, if a quantum system is initially represented by a wavefunction ψ that is *not* an eigenstate of the operator A , then after we observe the physical property A to be a , the quantum system must subsequently be represented by the eigenstate ψ_a . The formalism then seems to be telling us that a measurement on a quantum system alters the state of the system in a rather peculiar way: before we observe a quantum system there may be no definite value for a particular property, but when we make a measurement on that system we change the state of the system into one which has a definite value for the property we are measuring.

What I have attempted to present here is the bare fundamentals of the mathematical formalism of the theory of quantum mechanics. The formalism contains within it a very basic formula for mapping the behaviour of physical quantum systems onto the intuitive elements of our Newtonian picture of reality. However, as one can see already, the formalism also contains features which make for a rather counterintuitive picture of the behaviour of these quantum systems. The fundamentally probabilistic nature of quantum mechanics, the simultaneous indeterminacy of particular pairs of properties and the unusual change of state of measured systems combine to create a strange picture indeed. It is these strange features of quantum mechanics that are the source of the interpretational difficulties of the theory. Much effort was expended in the late 1920s in an attempt to develop an orthodoxy of interpretation that would alleviate these difficulties, and we will see over the course of this chapter the problems that still linger. Before we embark on this orthodoxy, however, I wish to say a few brief words about what we should expect an interpretation of quantum mechanics to achieve.

I should make it clear from the outset that the power of quantum mechanics is found in the fact that the formalism is very good at predicting the results of experiments. Of course, the formalism is only able to give probabilistic predictions about the behaviour of quantum systems, but these predictions do in fact agree with the statistical data that quantum experiments yield. Thus it is not unreasonable that we accept the theory of quantum mechanics as a good guide to constructing a quantum picture of reality.

I began this section with a claim that the abstract nature of the mathematical formalism of quantum mechanics does not lend itself to the sort of natural interpretation that is familiar in classical mechanics. Now that we are better acquainted with this formalism we can be more precise about this claim. The attractiveness of the Newtonian picture of reality is that there is a natural map from the mathematical formalism of classical mechanics onto our intuitive picture of reality, or the ‘manifest image’. For instance, we intuit that the world around us is composed of discernible individuals with determinate properties (position, momentum, *etc.*) and the mathematical formalism of classical mechanics represents this in an obvious manner. In contrast, there is nothing in the mathematical formalism of quantum mechanics that naturally suggests a physical picture of this sort. On the contrary, even with the limited formalism introduced above we can already see that the physical picture suggested at face value by the formalism is definitely not one composed of discernible individuals with determinate properties.

The initial interpretational debate loosely delineated itself along the lines of the following dichotomy: on the one hand, one simply takes the formalism at face value and then reads off the picture of reality that is suggested by the theory; on the other hand, one attempts to maintain some semblance of the Newtonian picture of reality and interpret the formalism of quantum mechanics accordingly. As it happens though, the former program has been guilty of unwittingly adopting many of the Newtonian intuitions also and, significantly for our present purposes, this includes the Newtonian picture of time. Thus it becomes apparent that the interpretational debate is a dispute over which of the Newtonian intuitions should be maintained, and of which we should dispose. What I ultimately wish to argue here is that we have no good reason to hold onto the Newtonian picture of time in the context of nonrelativistic quantum mechanics.

The very fact that there are multiple avenues open to the interpreter of quantum

mechanics is suggestive of an important feature of the sort of analysis in which we are engaging here. To a certain extent, a picture of reality derived from physical theory is underdetermined by the physical (observable and empirical) structures of that theory (recall the brief discussion of §2.4). For this reason, an analysis of the relative merits of differing interpretations of quantum mechanics will inevitably involve an examination of the package of metaphysical assumptions that each interpretation brings to the table (we will address this in more depth in the next chapter). These issues will become clearer as we continue to build our understanding of the interpretational debate in quantum mechanics. We now consider the interpretation of quantum mechanics that developed into somewhat of an orthodoxy: the Copenhagen interpretation.

4.4 The fifth Solvay conference, 1927

The fifth Solvay conference of 1927 was the setting for the beginning of the debate over the interpretation of quantum mechanics. In attendance were almost all of the contributors to the new quantum theory: Bohr, Heisenberg, Born, Pauli, Schrödinger, Dirac, de Broglie, Planck and Einstein among others. As it happened (Bacciagaluppi and Valentini, 2009), there was in fact no clear agreement concerning how the various quantum phenomena should be understood. However, through the ensuing decades the fifth Solvay conference came to be accepted as the place where an orthodox interpretation did emerge and this came to be known as the Copenhagen interpretation.⁷ Even though there are many (sometimes contradictory) positions that are often said to fall under the rubric of the Copenhagen interpretation, there is a clutch of core principles that can be taken to characterise the interpretation. Of the two general programs for interpreting quantum mechanics outlined above, the Copenhagen interpretation takes the mathematical formalism of quantum mechanics at face value and thus postulates no extra or no fewer entities than the mathematical description provides. Let us consider what this entails.

According to the Copenhagen interpretation the state of a quantum system is completely described by its wavefunction representation ψ ; there is nothing more concerning the state of a quantum system that could possibly be represented. How-

⁷In this context, Howard (2004) argues that the idea of a single orthodox Copenhagen interpretation is a myth invented by Heisenberg in the 1950s.

ever, despite this strict correspondence between the wavefunction and the state of the quantum system, the wavefunction is not supposed to provide a pictorial representation of its corresponding quantum system; it is not meant to represent a new kind of reality. The means by which the information contained in the wavefunction description produces a picture of reality is via the Born rule. Recall that, given an operator corresponding to an observable property of the system, the wavefunction decomposes into a linear combination of eigenstates with probability amplitudes as coefficients (4.1). The Born rule states that the square of the probability amplitude for each eigenstate represents the probability of finding the value of the observable property to be the corresponding eigenvalue. So while the wavefunction does not directly represent anything real, the Born rule provides the instructions for mapping the information provided by the wavefunction description into a physical picture representing the state of the quantum system.

Recall further the ramification which the Born rule has for wavefunction descriptions which are not eigenfunctions of the given operator: after operating on the original wavefunction ψ with an operator A , the resulting wavefunction must be an eigenstate of the operator, ψ_a . Thus according to the Copenhagen interpretation, before the measurement of the observable property A the state of the system is represented by ψ (which encodes all the probabilistic information about the possible results of the measurement), while after the measurement the state of the system is represented by $P_a\psi = \psi_a$ yielding the value a for the measurement with probability $|P_a|^2$. It is thus clear that due to the Born rule one major consequence of the Copenhagen interpretation is that it places considerable significance in the description of a quantum system on the act of observing or measuring that system. The wavefunction is said to ‘collapse’ on measurement from the original state to an eigenstate of the measurement operator. Significantly for our purposes in this chapter, this collapse is asymmetric in the sense that probabilistic information about the states that are not actualised is lost from the wavefunction description. Due to this loss of information, wavefunction collapse looks very different in one temporal direction than in the other: the Copenhagen interpretation of quantum mechanics is thus not time symmetric.

The final element of the Copenhagen interpretation of quantum mechanics is Bohr’s principle of complementarity. The principle of complementarity can be seen as growing out of an attempt to provide an account of how to picture the state of

a quantum system before any particular measurement is made. Imagine a quantum system represented by a wavefunction ψ that has two observable properties, A and B. The quantum system has certain propensities to collapse into one or other of the eigenstates of the operator A with a definite eigenvalue upon measurement of the observable property A. Likewise, the same quantum system has certain propensities to collapse into one or other of the eigenstates of the operator B with a definite (but possibly different) eigenvalue upon measurement of the observable property B. This is the case even when the observable properties A and B are forbidden by the quantum formalism to have simultaneously determinate values, i.e. the operators A and B do not commute. Thus the attained eigenvalues for observables A and B respectively cannot be representative of any definite properties of the system simultaneously. In such a case the physical properties are said to be *complementary*. The principle of complementarity then states that neither a description of the quantum system in terms of the properties associated with the measurement of observable A nor its description in terms the properties associated with the measurement of observable B are complete descriptions of the quantum system, but between them they form a complete, complementary description (Hughes, 1989, p. 228).

The combination of these ideas about the wavefunction description and the physical significance of the quantum formalism forms the core of the Copenhagen interpretation. The resulting quantum picture of reality is far removed from the picture of reality that arises from classical mechanics. According to the Copenhagen interpretation, it simply does not make any sense to talk about particular properties of a quantum system without reference to a particular physical situation. A quantum system might have one set of properties according to its wavefunction description with respect to some experimental setup and a different set of properties with respect to some other experimental setup. Since our only exposure to such a quantum system is through one or another experimental setup we must remain agnostic about the properties of the quantum system in the absence of a physical measurement. In such a situation we simply cannot ascribe any reality whatsoever to the physical properties of the quantum system. Although this is in some sense already embodied in Heisenberg's indeterminacy relation, the principle of complementarity implies that this indeterminacy not just a peculiarity of the measurement process but is an intrinsic feature of our interaction and description of reality. The principle of complementarity is then an epistemological principle about the conditions under

which we can have knowledge of a quantum system. By interpreting the wavefunction in these “epistemic” terms, wavefunction collapse is not thought to be a physical process in the Copenhagen interpretation and thus potential problems concerning nonlocal influence across a spatiotemporally bound wavefunction are avoided (though as we will see below, the Copenhagen interpretation, and indeed quantum mechanics itself, runs afoul of nonlocality elsewhere).

As mentioned above, this interpretation of quantum mechanics was not supported unequivocally at the fifth Solvay conference in 1927. The issue at the heart of this unrest over the Copenhagen interpretation was the lack of a definite reality ascribed to a quantum system. The main proponent of this view was Einstein who considered a physical theory that merely yielded information about probabilities an incomplete theory. In a sense Einstein believed a physical theory should be about physical reality and physical reality must have definite properties. In the present context this viewpoint can be thought of as exemplifying the second of the two general programs to which I had alluded above: namely that one could attempt to maintain some semblance of the intuitive worldview suggested by classical mechanics and interpret the formalism of quantum mechanics accordingly. Let us turn our attention to the argument against the Copenhagen interpretation.

4.5 The EPR argument

Einstein, Podolsky and Rosen (1935) (EPR) devised a thought experiment to challenge the physical picture of the Copenhagen interpretation. EPR were concerned that the Copenhagen interpretation did not provide a reasonable account of the physical reality of quantum systems. The project of the EPR paper can then be seen as the following: if one assumes a suitable criterion for the reality of a physical system, one can show using the quantum formalism that quantum mechanics is in fact an incomplete description of this reality. The argument is underwritten by the further assumption that every element of physical reality must have a counterpart in a complete physical theory. Before we consider the details of the EPR paper let us review in a little more depth the mathematical formalism behind the principle of complementarity.

Consider a quantum state represented by ψ and two operators A and B that are incompatible, i.e. there is no state ψ that is an eigenstate of both A and

B. Recall that the operation of A on ψ will result in an eigenstate of A with a corresponding eigenvalue a , $A\psi = a\psi_a$. If $\psi = \psi_a$ then the eigenvalue a will arise with certainty, but if this is not the case then a will arise with some probability less than one. By the same token, $B\psi = b\psi_b$ and the eigenvalue b will arise with certainty if $\psi = \psi_b$ but not otherwise. Since there is no eigenstate of A which is also an eigenstate of B (A and B are incompatible), the operation $B\psi_a$ then cannot yield an eigenvalue b with certainty. Therefore, knowledge of the value of the physically observable property A precludes knowledge of the value of the physically observable property B . Moreover, since ψ_a is not an eigenstate of B , attempting to obtain knowledge of the physically observable property B will inexorably alter the state of the system, collapsing it to say ψ_b , precluding any opportunity to establish with certainty the value of the observable property A thereafter. We are then left with the following disjunction: either the values of the observable properties corresponding to incompatible operators cannot exist simultaneously, or they can but they merely fail to be accounted for explicitly in the quantum mechanical description of such systems. It is this feature of quantum mechanics that is at the centre of the EPR argument.

EPR begin their challenge by stating the criterion of reality as follows: “[i]f, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity” (Einstein *et al.*, 1935, p. 777). So according to the criterion of reality, we can say that if the state of a quantum system is represented by a wavefunction ψ that is an eigenstate of some operator A corresponding to an observable property A , then there is an element of physical reality corresponding to the physical quantity a which is an eigenvalue of the operation of $A\psi$.

Consider now the behaviour of two quantum systems, L and R , described by the wavefunctions ψ_L and ψ_R , which are permitted to interact and are then subsequently separated, after which time it is assumed that any further interaction between the separated parts is not possible. Given appropriate interaction conditions, we can produce here what is called an *entangled* system: the state of the combined quantum system is treated as a single state $\psi_{L\&R}$ evolving in accordance with the Schrödinger equation. Given such a system in an entangled state, we are always able to find some pair of observables, A and B , with corresponding operators, A and B , such

that operating with A on the wavefunction representing the system allows us to make a definite prediction concerning the operation of B on that same wavefunction; this is known as the spectral decomposition theorem (Hughes, 1989). For entangled states whose spectral decomposition is not unique, we are further able to find multiple such pairs of operators that are pairwise incompatible: that is, two pairs A_1 and B_1 , A_2 and B_2 such that A_1 is incompatible with A_2 and B_1 is incompatible with B_2 .

Now consider $\psi_{L\&R}$ to be just such an entangled state, with two observable properties of system L corresponding to incompatible operators A_1 and A_2 that allow us to make definite predictions concerning the operation of B_1 and B_2 respectively on system R . If we make an observation of the property A_1 of system L , obtaining the physical quantity a_1 , we represent the resultant state as the collapsed wavefunction $P_{a_1}\psi_{L\&R}$. Since our original state has a non-unique spectral decomposition, knowledge of the physical quantity a_1 allows a definite prediction of the value of the observable property B_1 of system R . Making the further reasonable assumption that the principle of locality is not violated by this process, we can conclude that knowledge of this property of system R can be obtained *without disturbing that system*. In a similar fashion, if we make an observation of property A_2 of system L , obtaining the physical quantity a_2 (with the resultant state $P_{a_2}\psi_{L\&R}$), we can then make a definite prediction of the value of the observable property B_2 of system R , again without disturbing that system. Due to the criterion of reality both of the observable quantities of system R , relating to properties B_1 and B_2 , must then be elements of reality. Since B_1 and B_2 are incompatible operators, there are thus observable properties corresponding to incompatible operators which have simultaneous reality. When we submit this result to the disjunction above, we conclude that quantum mechanics is therefore not a complete description of quantum systems.

So what does all this mean for the interpretation of quantum mechanics? The Copenhagen interpretation which, recall, takes the quantum formalism to be a complete description of quantum systems, must reject the simultaneous reality of observable properties corresponding to incompatible operators and thus must reject the criterion of reality. The description of an observation of one of two entangled quantum systems would look very different according to the Copenhagen interpretation. The observable properties of the distant system would have no reality before the observation of the near system. The entangled systems would thus appear to remain connected after being separated in space, maintaining some direct influence

from one system to the other. Thus according to the Copenhagen interpretation, spatially separated elements of entangled systems appear to influence each other nonlocally. Moreover, the epistemic interpretation of the wavefunction, which, as we saw above, is able to avoid the nonlocality associated with wavefunction collapse by remaining agnostic about the reality of the quantum state, cannot be utilised to account for the nonlocal influence between entangled systems so long as this wavefunction description is taken to be complete.

In contrast, the EPR argument implicitly assumes that the principle of locality holds for all physical systems, even quantum ones. EPR claim that no “reasonable definition of reality could be expected to permit” the reality of the distant observable properties being dependent upon the observation on the near system. This is crucial to the EPR position and has been at the centre of the debate about how to interpret quantum phenomena. Furthermore, the criterion of reality suggests that we should expect observable properties of physical systems to be definite valued. In this way, the EPR position can be seen as an attempt to maintain these principles of the mechanistic worldview.

The EPR conclusion that quantum mechanics is an incomplete description of quantum systems also raises some interesting issues. If the wavefunction is not truly representative of the quantum state, what *would* a complete description be and how would it relate to the wavefunction description? One suggestion as to how to complete the wavefunction description is to insert the values of the physically observable properties of quantum systems, which are elements of reality according to EPR, into the quantum description as variables that are ‘hidden’ from within the quantum formalism. Such a proposal is called a *hidden variable* model of quantum mechanics. This puts a new spin on the Copenhagen interpretation of the wavefunction. The wavefunction remains representative of the possible knowledge one can gain about the state of a quantum system, but this knowledge is of a *real* quantum state and is epistemically limited to that part of the system that is not ‘hidden’.

We have considered here two different ways of interpreting the mathematical formalism of quantum mechanics. On the one hand we have the Copenhagen interpretation which ascribes no physical reality to the properties of a quantum system in the absence of a particular physical situation and which must permit nonlocal influences to account for the phenomena suggested by the formalism. On the other hand we have the hidden variable interpretation suggested by the EPR argument in which

the wavefunction description of a quantum system is an incomplete description but which maintains adherence to the principle of locality. During the time at which the debate between these two interpretations was still young, the experimental support for quantum theory was not rich enough to be able to shed much light on which interpretation might be the correct one. As is often the case in the philosophy of physics, the debate had to be argued on the grounds of scientific virtue and appeals to intuitive physical pictures. This situation was remedied in the 1960s through the work of John Bell.

4.6 Bell's theorem

Bell's (2004) seminal paper of 1964 grounds the thought experiment at the centre of the EPR argument in a physical quantum system. Whereas the EPR argument was concerned with the general mathematical formalism describing some entangled quantum system, Bell analyses the EPR argument in terms of a specific quantum system first recognised to be relevant to the EPR argument by Bohm and Aharonov (1957). I shall refer to this quantum system as the EPRB experiment. Bell takes the conclusion of the EPR paper seriously and formulates its central assumption, that “the result of a measurement on one system be unaffected by operations on a distant system with which it has interacted in the past” (Bell, 2004, p. 14), mathematically. What Bell then shows in his work is that no theory containing this assumption that makes predictions concerning quantum phenomena can be compatible with the statistical predictions of quantum mechanics. Thus, Bell's work establishes an experimental basis on which the conclusions of the EPR analysis of quantum mechanics can in fact be tested. In this section we will encounter the details of the EPRB experiment and Bell's theorem. Rather than pursue an exposition of Bell's formulation of his theorem, I shall instead characterise this theorem in terms of a more intuitive analysis due to Wigner (1970).

Let us begin by considering two particles in an entangled state called the *spin singlet* state. The spin singlet state can be thought of as the state of two particles whose spins must sum to zero, i.e. they have equal and opposite spin. For spin $\frac{1}{2}$ particles this is achieved when one particle has spin $+\frac{1}{2}$ (spin ‘up’, abbreviated by $+$) and the other has spin $-\frac{1}{2}$ (spin ‘down’, abbreviated by $-$). If the wave function is given by ψ , the exact representation of the spin singlet state in quantum mechanics

is the linear combination,

$$\psi = \frac{1}{\sqrt{2}}(|+, -\rangle - |-, +\rangle), \quad (4.2)$$

where the two terms of the expression represent a quantum state with spin up/spin down and spin down/spin up respectively for the two particles. The factor at the front is a normalisation factor.

Before continuing, a few general words about spin are in order. Spin is a binary property of a quantum particle which is defined in terms of a particular direction in three dimensional space. An intuitive way to imagine the connection between spin and direction is to imagine the axis of rotation of a rotating spherical particle. A particle which is rotating anticlockwise about the positive x direction will have spin up along the x axis. In terms of the quantum formalism, the process of observing the spin of a particle along the x -axis is represented by a spin operator S_x operating on the wavefunction ψ representing the quantum state. There are three such operators corresponding to the three spatial dimensions, S_x , S_y and S_z . As it turns out, these three operators are mutually incompatible operators. This is a large part of the reason why observations of spin values for quantum particles feature heavily as grounds for testing the EPR argument.

Wigner's account of Bell's theorem involves two spin $\frac{1}{2}$ particles produced in a spin singlet state; let the particles be denoted L and R as above. The two particles are then separated to distances such that measurements carried out on L are spacelike separated from measurements carried out on R , and such that no further interaction between L and R is possible. The system is then subject to two different measurements: a measurement of the spin component of particle L and a measurement of the spin component of particle R along any of three different spatial directions, ω_1 , ω_2 and ω_3 . There are nine possible ways to make these measurements, each yielding four possible results: both particles spin up, both spin down, one up and one down and *vice versa*. Since an appropriate measurement on one particle yields certainty about the properties of the other, and adopting the EPR criterion of reality, the results of all possible measurements must be elements of physical reality. These elements are then the hidden variables of the spin singlet state. Paramount to this analysis is the explicit use of an assumption of the EPR argument that the spin state of particle L is independent of the outcome of the measurement made

on particle R and *vice versa*. This can be seen as a restatement of our principle of locality in that any dependence would have to be mediated in a spatiotemporally contiguous way. Furthermore, it is assumed that the hidden variables are independent of which measurement is actually performed on either of the two particles. Let us call this assumption the principle of independence.

The spin singlet state dictates that along any particular direction the spin state of the two particles must be anti-correlated: that is, if one particle has spin up along a certain direction, the other must have spin down along the same direction and *vice versa*. Thus, if particle L was measured to have spin up in the ω_2 direction, and particle R was measured to have spin up in the ω_3 direction, then the spin singlet state determines that particle L has spin down in the ω_3 direction and particle R has spin down in the ω_2 direction. This restricts the space of possibilities for the measurement results and allows us to construct the possible combinations of spin states that the particles can have. There are eight such hidden states. Let the spin states of the two particles be characterised by the ordered tuple $(\sigma_1, \sigma_2, \sigma_3; \tau_1, \tau_2, \tau_3)$ where σ and τ represent the spin, + or -, of particles L and R respectively along directions ω_1, ω_2 and ω_3 . Using this notation, we can represent the eight possible combinations of spin values for the two particles, labelled as hidden variable states:

$$\begin{array}{ll}
 \lambda_1 : (+, +, +; -, -, -) & \lambda_5 : (-, +, +; +, -, -) \\
 \lambda_2 : (+, +, -; -, -, +) & \lambda_6 : (-, +, -; +, -, +) \\
 \lambda_3 : (+, -, +; -, +, -) & \lambda_7 : (-, -, +; +, +, -) \\
 \lambda_4 : (+, -, -; -, +, +) & \lambda_8 : (-, -, -; +, +, +)
 \end{array}$$

Each ordered tuple can be thought of as a particular state of the system which has a certain probability of occurring. We already have enough information to construct some properties of these probabilities. Firstly, the sum of all the probabilities must be one. In addition, any particular spin result for any particular direction occurs exactly half the time, i.e. in four hidden variable states. For example, spin down for particle L along direction ω_2 occurs in states $\lambda_3, \lambda_4, \lambda_7$ and λ_8 . Thus, the probability of any one particular spin result for either particle is one half. Furthermore, due to the anti-correlation properties of the spin singlet state that we have already encountered, the probability for the particles having alternate spins along the same direction is one and that for aligned spins is zero.

So far the properties of these probabilities are in agreement with quantum mechanics. To this degree, Wigner's model is a good model of quantum systems (insofar as quantum mechanics is a good model for quantum systems). However, it is possible to use the hidden states to create an inequality relating these probabilities through which the predictions of the Wigner model and those of quantum mechanics diverge. This divergence arises for measurements of different spin directions on each particle. Consider the following: inspection of the hidden states reveals that the probability of measuring L as spin up in the ω_1 direction and R as spin up in the ω_3 direction is the sum of the probabilities of only two of the possible eight spin combinations, $P(\lambda_2) + P(\lambda_4)$. Likewise, the probability of finding L as spin up in the ω_2 direction and R as spin up in the ω_3 direction is $P(\lambda_2) + P(\lambda_6)$ and the probability of finding L as spin up in the ω_1 direction and R as spin up in the ω_2 direction is $P(\lambda_3) + P(\lambda_4)$. If we accept that probabilities must always be positive, we can then create the following trivial inequality:

$$P(\lambda_2) + P(\lambda_4) \leq P(\lambda_2) + P(\lambda_6) + P(\lambda_3) + P(\lambda_4). \quad (4.3)$$

This all seems rather straightforward. However, according to quantum mechanics, the probability of finding spin up for two entangled particles along two different directions of measurement ω_i and ω_k is given by $\frac{1}{2} \sin^2 \frac{1}{2} \theta_{ik}$ where θ_{ik} is the angle between the directions. Substituting this into (4.3) yields

$$\frac{1}{2} \sin^2 \frac{1}{2} \theta_{31} \leq \frac{1}{2} \sin^2 \frac{1}{2} \theta_{23} + \frac{1}{2} \sin^2 \frac{1}{2} \theta_{12}. \quad (4.4)$$

It is not immediately clear that this inequality holds. We can explore the behaviour of this inequality by setting $\theta_{12} = \theta_{23} = \frac{1}{2} \theta_{31}$. Substituting these values into (4.4) gives

$$\frac{1}{2} \sin^2 \frac{1}{2} \theta_{31} = \frac{1}{2} \cdot 4 \sin^2 \frac{1}{2} \theta_{12} \cdot \cos^2 \frac{1}{2} \theta_{12} \leq \sin^2 \frac{1}{2} \theta_{12}.$$

Solving for θ_{12} ,

$$\begin{aligned} \cos^2 \frac{1}{2} \theta_{12} &\leq \frac{1}{2}, \\ \theta_{12} &\geq \frac{\pi}{2}. \end{aligned}$$

Using the above relation between the angles, one can see that (4.4) is *violated* when-

ever $\theta_{31} < \pi$. That (4.4) can be violated at all should raise some eyebrows; (4.3) is trivially constructed and (4.4) is derived therefrom. The problems occur when we consider measurements of different directions for each particle and obtain results of aligned spin. Let us consider what exactly the predictions are for both the Wigner model here and orthodox quantum mechanics in these cases.

By inspecting the set of hidden states, it can be shown that the probability of measuring aligned spins along different directions for each particle is $\frac{1}{3}$. If each of the eight hidden states is just as likely as any other to be actual, of the nine possible measurement setting combinations $((\sigma_i; \tau_k)$ for $i, k = 1, 2, 3$), λ_1 and λ_8 have zero possibilities with aligned spins in different directions and $\lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$ and λ_7 have four possibilities each. Thus $\frac{6 \times 4}{8 \times 9} = \frac{1}{3}$. The predictions of quantum mechanics, however, are vastly different. We have seen as such in (4.4): the probability for measuring aligned spins along any directions of measurement ω_i and ω_k is given by $\frac{1}{2} \sin^2 \frac{1}{2} \theta_{ik}$ where θ_{ik} is the angle between the directions. Hence, quantum mechanics predicts a particular type of correlation between the two particles in the EPRB experiment: the probabilities of the outcomes of measurements on particle L are dependent upon the measurement settings employed to measure particle R , and *vice versa*. In other words, the probabilities of the outcomes of measurements on particles L and R cannot be factorised into a product of independent probabilities; this is often referred to as a violation of the *factorisability* condition.⁸ Bell's theorem can then be stated in the following terms: no theory that satisfies the factorisability condition can reproduce the predictions of quantum mechanics.

The significance of this result should not be underestimated. Recall the origin of this analysis: there existed a disagreement as to how the mathematical formalism of quantum mechanics should best be interpreted. On the one hand, the Copenhagen interpretation took the formalism at face value which resulted in a picture of reality in which quantum mechanics violated the principle of locality. On the other hand, when the principle of locality was maintained as an assumption, the conclusion that was drawn was that quantum mechanics was not a complete description of atomic systems. With the introduction of Bell's theorem, we now have a formula for testing which of these positions is closer to a description of quantum systems. If one were to carry out an EPRB experiment in which the angle between the measurement directions ω_3 and ω_1 was less than π and the measurement direction ω_2 was such that

⁸See, for instance, Butterfield (1992).

it bisected the other two directions, then one would be able to collect experimental results that would confirm at most one of these two descriptions of quantum systems.

Let me build the suspense some more. If the probability of measuring aligned spins along different directions for each particle of an EPRB experiment turned out to be $\frac{1}{3}$, this would provide quite damaging evidence against quantum mechanics being a complete theory. Physicists would then have to go back to the drawing board to develop a supplementary theory describing the behaviour of atomic systems, some of the more intuitive principles of the mechanistic worldview would be salvaged and the Newtonian picture of reality would be reinvigorated. If, however, the probability of measuring aligned spins along different directions for each particle of an EPRB experiment turned out to be dependent upon the angle between the measurement settings of the apparatus in the way predicted above, quantum mechanics would be vindicated and Wigner's toy model would be falsified. This is no mere formality though; Bell's theorem implies that if quantum systems actually behaved according to these statistics then no theory satisfying the factorisability condition could provide an empirically adequate description. It would then be the case that the result of a measurement of one system *would* be affected by operations on a distant system with which it has interacted in the past; the principle of locality itself would be shown to be a misguided principle about the nature of our reality. Either way, something profound about the picture we have of reality is to be discovered from the EPRB experiment.

A series of experiments were in fact carried out by Aspect, Delibard and Roger (1982a) and Aspect, Grangier and Roger (1982b) which confirmed the predictions of quantum mechanics: the probability of measuring aligned spins along suitably chosen different directions for each particle of an EPRB experiment *is* dependent on the angle between the measurement settings of the apparatus in exactly the way predicted by quantum mechanics. The two particles (or photons, as was actually the case in the experiments) of the EPRB experiment really are correlated in this way and hence the factorisability condition is violated by actual quantum systems. The significance of this result is that it shows that causal influences do in fact extend across spacelike separated spatiotemporal distances; the principle of locality is unconditionally violated by the theory of quantum mechanics! With this result many physicists merely accepted the lessons of the Copenhagen interpretation of quantum mechanics: one cannot talk of a determinate reality for a physical system

before an observation has been made and one must accept that the universe permits nonlocal influences between quantum systems, and thus that there is a fundamental incompatibility between quantum mechanics and the theory of relativity. This is not, however, the exclusive moral that can be drawn from this result.

4.7 EPRB and retrocausality

While the Aspect *et al.* experiments amount to strong evidence that nonlocal influences are something that cannot be avoided in quantum theory⁹, this does not necessarily imply that quantum mechanics is incompatible with the theory of relativity. The crucial point to note is that the theory of relativity is not incompatible with nonlocality *per se* but rather with a violation of what I will label here *action-at-a-distance*: that is, some sort of causal influence that has propagated without spatiotemporal contiguity. We can set this opposed to *action-by-contact* which we can think of as causal influence that propagates along timelike curves. What I will set out here is a scheme by which we can strictly maintain action-by-contact, as it were, and still explain the nonlocal character of quantum mechanics.

One issue which I have deliberately skirted around till this point is the generality of Bell's 'general' analysis of the hidden variable scheme. Recall that we explicitly assumed two principles concerning the hidden variable states: the spin state of particle *L* is probabilistically independent of the outcome of the measurement made on particle *R* and *vice versa*; and the hidden variables are probabilistically independent of which measurement is actually performed on either particle *L* or particle *R*. We have called these the principle of locality and the principle of independence respectively. Bell obeys these principles in his general analysis of the hidden variable scheme, but it may be the case that a hidden variable model not encompassed by Bell's analysis could be constructed, by violating one or both of these principles, which could recover the predictions of quantum mechanics despite Bell's theorem. One option available to us would be to construct a hidden variable model that explicitly violated the principle of locality. A model such as this had in fact been developed well before Bell's theorem by Bohm (1952a). Our intention here, however, is to recapture the notion of action-by-contact rather than explicitly jettison

⁹Although, see Christian (2007) for an apparently exact, locally causal model of the EPRB correlations.

the principle of locality. Thus let us then pursue the other option available to us, which is to violate the principle of independence.

The principle of independence asserts that the hidden variable state of a quantum system is independent of the type of measurement that is to be performed on that system. A violation of this principle would imply that the hidden variable state of the system depends directly upon the *future* measurement settings. Rudimentarily this amounts to the following claim: if we perform the EPRB experiment on some pair of particles in some hidden state λ_i and measure, say, ω_2 on particle L , then *had* we made a different measurement, say ω_1 on particle L , then the hidden state of the particles would also have been different, say λ_k . Notice the counterfactual nature of this claim; if we were to adopt a counterfactual theory of causality (more on this §5.2.2), we would then be at liberty to say that the measurement settings of the EPRB apparatus *causally* influence the hidden variable state of the quantum system. Since the hidden variable state of the two particles temporally precedes the eventual measurement, we say that this relationship is *retrocausal*. Thus postulating a violation of the principle of independence to circumvent Bell's analysis of hidden variable models of quantum mechanics is akin to hypothesising retrocausal influences in quantum systems.

To see that this will help us explain the strange correlations of EPRB type experiments, consider the causal relationship between the setting of the device measuring particle L and the hidden variable state of the two particle system: the spin state of the distant particle R is trivially dependent on the hidden variable state; as a consequence, particle R is in turn dependent on the setting of the device measuring particle L ; and this, of course, is just the nonlocal correlation between the particles predicted by quantum mechanics. Thus we should be able to recapture the predictions of quantum mechanics despite the lessons of Bell's theorem by giving up the principle of independence. Let us look at the mechanics of this possibility in more detail.

If the setting on the left hand side of the experiment is ω_1 and the setting on the right hand side of the experiment is ω_3 , then the pair of particles will have a hidden variable state which reflects this situation, i.e. the hidden variable state will contain a measurement outcome for direction ω_1 on the left and ω_3 on the right. Let us represent such a hidden state as $(\sigma_1; \tau_3)$ where the index represents the direction of measurement and both σ and τ can take values of $+$ or $-$ as before. More generally,

we can represent all the hidden variable states, λ_j , by the pair $(\sigma_i; \tau_k)$, where i and k can take the values 1, 2 and 3 corresponding to the measurement directions ω_1 , ω_2 and ω_3 , and j enumerates all the possible combinations of hidden variable states (there are in fact 30 such states considering that the spins of the particles cannot be correlated along the same direction of measurement).

One important feature of such a retrocausal account of the EPRB experiment is that given the hidden variable state corresponding to a future pair of measurements along two particular directions, the remaining spin states for the other directions of measurement simply do not play a part in forming the hidden variable state for the entangled pair. If the measurement apparatus is set to measure the spin of the particle along direction ω_3 on the left and ω_2 on the right, no spin state corresponding to ω_1 and ω_2 on the left, nor ω_1 and ω_3 on the right need partake in the formation of the hidden variable state. No longer must we attribute simultaneous reality to all the possible states of the system corresponding to all the possible future measurements we could make. Only the *actual* future measurements of the system are accounted for in the hidden variable state. Moreover, there is no loss of probabilistic information associated with the measurement process as in the Copenhagen interpretation and thus there is a certain temporal symmetry in the evolution of quantum systems. This feature also avoids a potential problem of the Bell formalism that the hidden variable probabilities are probabilities over incompatible observables.

The correlations predicted by quantum mechanics are rather specific. For any measurement settings ω_i and ω_k , the probability of finding correlated spins is given by $\sin^2 \frac{1}{2}\theta_{ik}$ and the probability of finding anti-correlated spins is given by $\cos^2 \frac{1}{2}\theta_{ik}$, where θ_{ik} is the angle between the directions. The hidden variable states $(\sigma_i; \tau_k)$, unlike the Wigner hidden variable states λ_1 through λ_8 , are not naturally attributed equal probability and thus it is a simple task to ascribe to these states the exact probabilities predicted by quantum mechanics.

The important feature of this physical picture of a retrocausal EPRB experiment is that even though particles L and R influence each other across a spacelike separation, this influence is mediated by two timelike worldlines: one from the preparation of the system to the measurement of particle L ; and the other from the preparation of the system to the measurement of particle R . Thus the nonlocal influence is explained by two causally contiguous worldlines that ‘zigzag’ backwards and forwards in time. Even though no spatiotemporally contiguous process could *ordinarily* me-

diate this influence, now that we have postulated retrocausal influences in quantum systems we are able to provide a description of this nonlocal process that does maintain spatiotemporal contiguity, and is thus an action-by-contact process. There may be a nonlocal causal influence, but there is no action-at-a-distance.

4.8 The lesser-of-two-evils

The hypothesis of retrocausal processes in quantum mechanics is not a new idea: Costa de Beauregard (1953) first floated the idea of retrocausality with EPR type situations in mind.¹⁰ However, it has not been a very popular idea. One surmises that the major reason for this is that the positing of backwards-in-time causal influence does not fit particularly well with the Newtonian picture of reality; in particular, the generative picture of determination, which is a defining element of the Newtonian picture of time. This seems a rather peculiar argument when one considers that the most popular alternative to retrocausality, reading Bell's theorem as confirming the existence of nonlocal causal influences in quantum systems, does not fit particularly well with the relativistic picture of a causally contiguous (i.e. local) spacetime.¹¹ (This peculiarity is compounded when one also considers that both the Lagrangian picture of determination, where the initial and final states of a system equally determine dynamical behaviour, and general relativity, where solutions to the Einstein field equations are equivalence classes of diffeomorphism invariant four dimensional spacetimes, obviously defy this Newtonian schema; but more on this in §6.8 after we have a better grasp on the nature of causality.) If both of these responses to Bell's theorem are bound to jettison certain elements of the Newtonian picture and maintain certain others, how are we to decide which is 'less evil' with respect to the task of procuring a coherent picture of reality?

We are already in a position to look favourably on retrocausality in light of the fact that the part of the Newtonian picture with which it conflicts, generative determination, arises largely in the context of 18th century metaphysics, while the part of the Newtonian picture with which nonlocality conflicts is a central tenet of

¹⁰See also Argaman (2008), Costa de Beauregard (1976, 1977), Cramer (1980, 1986), Hokkyo (1988), Miller (1996, 1997), Price (1984, 1994, 1996, 1997, 2001, 2008, 2010), Rietdijk (1978), Sutherland (1983, 1998, 2008) and Wharton (2007, 2010).

¹¹The many-worlds interpretation is another popular alternative that purports to provide a local interpretation of quantum mechanics.

relativity theory. Granted, we should not discard what has thus far been an integral part of our picture of reality simply because we might think it *passé*; but by the same token we should not discard a potential solution to the interpretational difficulties of quantum theory simply because it challenges our Newtonian conception of reality. There are further question marks that need to be addressed, however, before a significant challenge can be mounted against this old guard (such as the apparent paradoxes that could arise in a retrocausal context). The task remains then to construct a retrocausal conception of reality with which to mount this challenge, and this is the aim of the next chapter: I examine the physical possibility of retrocausality in a reality such as ours.

Retrocausality at No Extra Cost

We have seen how the proposal of retrocausality is able to alleviate some of the interpretational difficulties posed by Bell's theorem: by permitting retrocausal influences in quantum mechanics, the strange correlations of the EPRB experiment can be accounted for in an action-by-contact manner. The main element of the Newtonian picture of reality that conflicts with this proposal is the generative picture of determination, or the Newtonian picture of time. Maudlin (2002) provides a particularly clear expression of this conflict when he argues that retrocausality is fundamentally at odds with the “metaphysical picture of the past generating the future” and thus cannot be entertained as a metaphysical possibility in a reality such as ours. We will return to assess the plausibility of Maudlin's argument in Chapter 6. In this chapter, however, I build an independent case in favour of retrocausality as a physical possibility as a counterbalance to Maudlin's picture.

5.1 Introduction

The metaphysics of retrocausality is often broached in the philosophical literature in and around discussions of time travel and causal paradoxes and there seems to be a general sentiment that there is nothing *manifestly self-contradictory* about the idea, strange though it may seem at first. The purpose of this chapter is to develop these metaphysical considerations into a carefully considered picture that coheres with the possibility of retrocausality and that is not precluded by our current physical theories.

The picture developed in this chapter is a conglomeration of developed ideas collected together from various contexts. The goal here is to combine these ideas into a single coherent picture which will assist us in forestalling some perceived meta-

physical problems with retrocausality. I begin by setting out in §5.2 two relatively uncontroversial positions that will serve as a solid conceptual foundation upon which to develop our retrocausal picture: the block universe model of time in §5.2.1 and the interventionist account of causation in §5.2.2. There are then two metaphysical intuitions that must be dismantled. The first is our ordinary asymmetric causal intuition: in §5.3.1 I describe an argument from temporal symmetry against the plausibility of extending our asymmetric causal intuitions to the microscopic realm. The second is our ordinary intuition about epistemic access to the past: in §5.3.2 I present an argument that clears a logical space for retrocausality at the expense of our intuition that our past is necessarily epistemically accessible independent of our own future actions. The claim here is that quantum mechanics is a theory that occupies this logical space. This then clears the way to build a symmetric picture of causation: in §5.4 we sequester a model of agent deliberation that permits us to strike a harmony between our causal intuitions, such as free will and unidirectional causation, and the picture derived from spacetime physics that future events are fixed within a deterministic and causally symmetric framework.

One way to imagine the choice between the competing interpretational schema of the last chapter is in terms of *ideological economy*.¹ According to Quine (1951), the ideas that can be expressed in a theory comprise the ideology of the theory. The ideological economy of a theory is then a measure of the economy of primitive undefined statements employed to reproduce this ideology; fewer primitive statements imply a more economical ideology. One argument that is often made against retrocausality is that introducing retrocausal influences into the quantum mechanical ideology is less economical than rejecting local hidden variables. The goal of this chapter is to show that introducing retrocausality incurs no economical cost because the ingredients required to build a retrocausal quantum picture of reality are given to us for free by the metaphysical structure of our existing physical theories and the epistemological structure of our experiences.

¹This terminology is derived from Quine's (1951) distinction between the *ontology* and the *ideology* of a theory.

5.2 Foundations

5.2.1 The block universe model of time

Before introducing the account of causation that we will utilise throughout this examination, let me say a brief word about the metaphysical position I will be taking with regards to time. I adopt here a temporal model popular among many physicists and philosophers: the block universe model of time (recall §2.1.1). Rather than modelling reality as a three dimensional space evolving under the passage of time, reality is envisaged according to this view as a four dimensional block of which time is a mere passive ingredient. The safety in adopting this stance on time is the compatibility we gain with contemporary theories of spacetime; this was discussed in depth in Chapter 2. There is a definite advantage for adopting such a temporal model. The power of the block universe model, if considered in the right way, is that it wears its spatiotemporal consistency constraints on its sleeve, so to say. We are about to embark on an analysis of models of physical systems that extend across both space and time, and considering these as four dimensional static systems will be especially beneficial to evaluating their spatiotemporal consistency. Hopefully we will soon see that both the account of causation introduced below and the block universe model of time are naturally suited to describing the sorts of physical systems and processes under consideration in the remainder of this thesis.

5.2.2 The interventionist account of causation

The account of causation that I adopt is the interventionist account of causation, as introduced and defended by Woodward (2003). The essential ingredient in this account is the notion of *manipulability* or *control*: according to this account, we say that C is a cause of E just in case there is some possible intervention that can be carried out on C that will change E in some way or other, holding fixed all other properties of the system containing C and E . Woodward's account is explicitly counterfactual in the sense that there need be only some *possible* intervention that can be made on C to bring about a change in E . The advantage of this account is that it can be employed to provide causal explanations without requiring that there exists a complete description of some spatiotemporal process connecting C and E ; we will use this feature explicitly below. To understand this account of causation

more clearly, let us consider an illustrative example.

Imagine the ignition system of a car. It seems that we would want to say that the turning of the key in the ignition (event K) is the cause of the starting of the car's engine (event E). According to the interventionist account we can say that K is indeed the cause of E since it is possible to carry out an intervention on K , by not turning the key say, that will change E in some way or other, in this case the engine would simply not start, provided all the other elements contributing to the system were held fixed. We can in fact claim a causal connection here without explicitly spelling out the mechanism by which the turning of the key brought about the starting of the engine. However, this does not mean that we cannot spell out such a mechanism if we wished.

Consider the mechanical chain of events connecting the turning of the key to the starting of the engine: turning the key (event K) completes the circuit between the car's battery and the starter motor (event C) which then starts the starter motor spinning (event S); the spinning starter motor then turns over the drive shaft of the engine (event D) which starts the pistons drawing in and then combusting the fuel (event P); the combusting fuel powers the engine to start running (event E). We have a chain of events, $K \rightarrow C \rightarrow S \rightarrow D \rightarrow P \rightarrow E$, with a mechanical account of how each event brings about the next. However, the content of any causal claim about any two of these events according to the interventionist account of causation is not that there exists a mechanical connection between the events. The key to the interventionist account is to imagine that each of these events is a handle or variable that can be manipulated and controlled. Accordingly, what makes each event the cause of the next is the fact that there exists a functional dependency between the variables; that is, some possible intervention on a particular variable will (over a range of conditions) bring about a consistent change in the values of the variables further down the chain. If we were to intervene on the above system by replacing the battery with an old or faulty battery, the starter motor would fail to spin, thus changing the value of the variable associated with event S (on or off, say) from what it would have been had we not made the intervention.

The chain of events may be more complex than the example above; events might have multiple causes or multiple effects. We can extend our example by imagining that event K , the turning of the key, also completes the circuit between the car battery and the dashboard, lighting up all the instruments inside the cabin (event

I). We could also imagine that for the drive shaft to turn over (event D), the car's gearbox must be disengaged (event G). We can establish which correlations are causal by imagining possible interventions of these new variables while holding the rest of our variables fixed. If we sever the connection between the car battery and the dashboard, the battery would still connect to the starter motor bringing about the ignition of the car's engine. Thus, the dashboard lighting up is not a cause of the car's engine starting, even though these events are very often correlated. However, if we engage the gearbox with the drive shaft and attempt to start the car, we find that the car stalls. Whether the gearbox is engaged or not, i.e. whether the variable associated with event G is one value or another, has a functional relationship to whether the car starts or not and is thus a cause of event E .

It is the adoption of this account of causation that permits us to talk about retrocausal influences in the EPRB experiment. We characterised the violation of independence in §4.7 as implying that the hidden variable state of some pair of particles prior to some measurement would have been different had the measurement been different. This counterfactual relationship qualifies as a causal relationship according to the interventionist account of causation.

There is one further issue which arises from this account of causation that will be crucial to our characterisation of retrocausality later in this chapter. It will be beneficial for our purposes here to view the interventionist account of causation as a kind of genealogical account of how we, as agents, come to acquire the concept of causation in cases where we have no possibility of intervening in the world around us. To begin demonstrating how this might be the case, consider the way we might employ causal concepts to describe a situation in which it is impossible for us as humans to intervene on the system. The gravitational pull of the moon is responsible for the ebb and flow of the tides, and we would want to say that the moon *causes* the tides' ebb and flow. Even though it is implausible for us to actually manipulate and control this system, we can attribute our causal intuitions in this sort of case to an ability to extend our causal intuitions from cases in which we can manipulate and control. Through our knowledge of the gravitational interaction between the moon and the tides, we can predict with confidence what the effect of some imagined (but perhaps physically impossible) intervention would be if we could in fact bring it about. It is this sort of knowledge which we usually gain by physical intervention and experimentation that allows us to make claims about the causal relations that exist

within a system. Thus, it seems reasonable that we extend these causal notions to cases in which we do not in fact have the requisite ability to manipulate and control.

I mention this feature of the interventionist account here to highlight the fact that a consequence of this view is that our role as agents in the world can be seen as at the root of our concept of causation. We will take this idea up again below where we will be in a position to expand on it in more depth. For now I simply wish tentatively to broach the outcome of this genealogical sketch that a being who interacted with the world differently to us as an agent would have a very different concept of causation to the one that we have.

With these metaphysical foundations in mind, let us now move on to dismantling two of our ordinary temporal intuitions.

5.3 Dismantling intuitions

5.3.1 Macroscopic intuitions, microscopic symmetry

A familiar intuition, indeed one that seems almost trivial, is that the properties of interacting systems are independent before they interact. This is built upon the observation of many instances where this apparent principle holds true. In macroscopic systems we take this principle for granted. However, Price (1996, 1997) asks the question whether we are justified in extrapolating this familiar macroscopic principle to considerations of microscopic systems. Let us consider Price's analysis.

Firstly, it seems that the origin of this principle is related to the asymmetry of thermodynamics. When systems evolve from states of disequilibrium (lower entropy) to states of equilibrium (higher entropy) it is because the initial conditions are special; namely, the initial conditions are low entropy. Thus, if we were to consider a macroscopic system evolving in the reverse temporal direction, it would look strange because it would appear that highly correlated *incoming* influences were converging from disparate regions of space (imagine a pile of rubble 'un-collapsing' into a building) and these would be associated with a decrease in entropy. In such a case the violation of the principle that physical processes are uncorrelated before they interact would be a direct product of the violation of the second law of thermodynamics. Looking forwards in time again we can see that the temporal asymmetry which manifests itself in the correlations between *outgoing* influences is a result of

special (low entropy) initial conditions and not a result of an inherent asymmetry within the laws of physics.

It appears to be assumed that this principle of outgoing correlations but incoming independence holds in the microscopic case just like in the macroscopic case. However, explaining this temporal asymmetry of microscopic systems in terms of boundary conditions simply does not work. The boundary conditions explanation is based upon the temporal asymmetry of entropy change. In a microscopic system, such as that of two particles which come together, interact and then separate, there is no entropy gradient of the sort we find in the macroscopic case to indicate a temporal orientation to the interaction. The temporal reverse of the interaction would look much the same as in the ordinary temporal direction; this is a function of the temporal symmetry of the dynamical laws of the system. Thus, there seems to be no reason to assume that outgoing correlations exist in one direction and not in the other. Furthermore, unlike in the macroscopic case, there is no *observed* asymmetry in microscopic systems that needs to be explained. We simply do not observe the independence of incoming particles nor the correlation of outgoing particles, yet we still assume that this principle holds for microscopic systems despite its incompatibility with the temporally symmetric nature of the dynamical laws of physical systems.

Therefore we are left with a dichotomy between two physical principles at the microscopic level: temporal symmetry in the dynamical laws of physical systems on the one hand and, on the other hand, the asymmetry of the independence of microscopic systems prior to interacting with each other. As such, one could make quite a persuasive argument against the independence of microscopic systems prior to interaction purely on symmetry grounds. Moreover, with no existing observational evidence in favour of the independence of incoming particles or the correlation of outgoing particles, it seems that such a principle may not deserve the status it currently enjoys in considerations of microscopic systems. If this principle is then abandoned, one is led to the conclusion that temporally symmetric causation in microscopic systems cannot be ruled out on analytic grounds. Thus if we take these considerations seriously then the nature of the physics in this case does not preclude a picture of reality that coheres with the possibility of retrocausal influences.

5.3.2 The bilking argument

In our normal conception of causation, causes precede their effects. A causally symmetric viewpoint opens up the possibility that effects can precede their causes. This, however, immediately creates some potential conceptual difficulties. To demonstrate these difficulties, let us imagine a pair of events which we believe to be causally connected: a cause, C , and an effect, E . Let us further imagine that this connection is retrocausal; E occurs earlier in time than C . On first appearances it would then seem possible to devise an experiment which could confirm whether our belief in the causal connection is correct or not. Namely, once we had observed that E had occurred, we could then set about ensuring that C does not occur, thereby breaking any retrocausal connection that could have existed between them. If we were successful in doing this, then we would have *bilked* the effect of its cause. This is the bilking argument.

The bilking argument seems to drive one towards the claim that any belief an agent might hold in the positive correlation between event C and event E is simply false. If this were the case then the agent would have to give up any belief in retrocausal influences between C and E . Dummett (1964) disputes that giving up this belief is the only solution to the bilking argument. In exploring the terms under which a belief in retrocausation can be maintained, Dummett suggests that what the bilking argument actually shows is that a set of *three* conditions concerning the two events, and the agent's relationship to them, is incoherent. In any incoherent set of conditions, all three conditions cannot hold simultaneously. Thus, depending on which of these three conditions fails to hold, there may be scope for an agent to maintain a belief that the later cause retrocausally influences the earlier event. To motivate these conditions, let us consider Dummett's own example.

Dummett imagines a tribe to exist with the custom of sending young men on a lion hunt to prove their bravery. The men travel for two days, hunt for two days and spend two days on their return journey. Observers travel with the young men and report back to the chief of the tribe whether the men acquitted themselves with bravery or not. While the young men are away, the chief performs dances intended to cause the young men to act bravely. Significantly, he performs these dances for the whole six days, i.e. for two days during which the events that the dancing is supposed to influence have already taken place. The chief notices that on

occasions when he dances, he subsequently learns that the young men had hunted bravely and, on occasions when he does not dance, he subsequently learns that the young men had hunted in a cowardly fashion. The chief thus observes there to be a positive correlation between his dancing and the young men's bravery and therefore maintains a belief in retrocausation.

Imagine further that we are to convince the chief that this practice of his were absurd. We arrange that the observers who had accompanied the hunt return early and report to the chief whether or not the young men had acted bravely. We then set a bilking challenge to the chief to dance if and only if the young men had *not* acted bravely. There are two possible outcomes of this challenge. If the chief accepts this challenge and dances then he must concede that his dancing does not ensure the bravery of the young men. Alternatively, imagine that the chief accepts the challenge and then discovers he is inexplicably unable to dance, i.e. his limbs will simply not move. Then the chief would have to admit that dancing is not an action which is within his power to perform. If this were to occur, however, it would then be fair to say that it is not the chief's dancing that causes the young men to be brave, rather it is the young men's bravery that makes possible his dancing. Thus, regardless of whether the chief dances or not, it seems that the chief must give up his belief in retrocausality.

It appears then that there are two incompatible conditions here concerning the chief's dancing: (i) there is a positive correlation between the chiefs dancing and the bravery of the young men; and (ii) dancing is within the power of the chief to perform. If the first condition is to hold, then the second condition must fail, and *vice versa*, as we have just seen. Dummett, however, suggests that an implicit third condition can be violated which allows both of these conditions to hold simultaneously and thus allows the chief to maintain his belief in retrocausality. To see this, let us first consider an agent who believes a certain action is effective in bringing about a *subsequent* event. Such an agent would believe the action to be the cause of the later effect. Dummett recognises that there is a connection between the foreknowledge the agent possesses about the subsequent event and the intention the agent has to perform the action. The agent only knows an event to occur in the future because they intend to bring it about by performing a certain action: the agent possesses knowledge in intention. This is in contrast to knowledge of the past which we can possess in more forms than merely in intention.

Let us then return to our example and imagine for the sake of argument that there is a parallel between the knowledge that the chief can possess concerning the bravery of the young men and the case of foreknowledge described here, i.e. the chief only knows that the young men are brave due to his intention to dance. This would then make our bilking challenge inconclusive. Since we can no longer arrange that the observers report the behaviour of the young men to the chief, we can no longer force the occurrence of a negative correlation. If we further rule out that there are no inexplicable incidents when the chief is unable to dance, then we are left with the original situation whereby the chief merely observes a positive correlation between his dancing and the young men's bravery and the chief can thus maintain his belief in retrocausation. To arrive at this result we have had to jettison the following condition: (iii) the chief has epistemic access to the behaviour of the young men independently of his intention to dance. These three conditions form a set which is shown to be inconsistent by the bilking argument.

Let us state these conditions in the more general terms we encountered at the beginning of this section.

- (i) There exists a positive correlation between an event C and an event E .
- (ii) Event C is within the power of an agent to perform.
- (iii) The agent has epistemic access to the occurrence of event E independently of any intention to bring it about.

An interesting point to notice at this stage is that these conditions do not specify in which order events C and E occur. If we consider why it is not the case that it is possible to bilk *future* effects of their causes, this is because condition (iii) fails to hold for future events. If knowledge about future events could be obtained independently of an agent's intention to perform certain actions, then it would be possible to bilk those future events of their causes; this would amount, in a way, to changing the events we already know to occur in the future. Since this sort of foreknowledge is not possible, we can consistently believe our actions to *bring about* the future. Conversely, if it were the case that some past event was known only through our intention to perform a certain action, then it would be consistent to believe our actions to *bring about* the past.

The conditions under which it is possible to maintain a belief in retrocausation are especially relevant to quantum mechanics. In fact, once we make a suitable specification of how condition (iii) can be violated, we find that there exists a strong symmetry between the conditions which need to hold to justify a belief in bringing about the past and what we find to be the case in quantum mechanics. Following the prescription of Price (1996, p. 174), let us not suppose that a violation of condition (iii) entails that the relevant agent has *no* epistemic access to the relevant past events independently of any intention to bring them about, rather let us suppose that the means by which knowledge of these past events is gathered breaks the claimed correlation between the agent's action and those past events. We can state our new condition as follows:

- (iv) The agent can gain epistemic access to the occurrence of event E independently of any intention to bring it about and without altering event E from what it would have been had no epistemic access been gained.

In the dancing chief example a violation of this condition would entail that every time the chief attempted to discover the behaviour of the young men he subsequently affected their behaviour to be different from what it would have been had he not attempted his discovery. In those cases where the chief makes no attempt to discover the behaviour of the young men, we are back to our original violation of condition (iii).

The nature of this weakened violation of condition (iii) should look familiar; it is just the sort of condition we would expect to hold if the system in question were a quantum system. To be more explicit, let me take us on a brief detour and construct an example of a simple quantum system to see that this is the case. Imagine a quantum system is prepared to be in a state ψ at time t_0 and at time t_1 the system is to be measured. In the orthodox interpretation of quantum mechanics the wavefunction representing the system evolves according to the Schrödinger equation from t_0 until the time of measurement at t_1 wherein the wavefunction collapses to one or other of the eigenstates of the operator associated with the measurement.

Let us now imagine that we are an agent who believes in a retrocausal influence from the measurement at t_1 on the state of the system after preparation at t_0 . To begin with, we cannot be subscribers to the Copenhagen interpretation because this belief is incompatible with the belief that the wavefunction description

is a complete description of the quantum state. Due to the asymmetric nature of the measurement process and the Born rule, there is simply no correlation at all between the wavefunction after t_0 and the measurement at t_1 . To make a claim that there is a correlation between the state of the quantum system after t_0 and the measurement at t_1 , we cannot take the wavefunction to be a complete representation of the quantum system. Thus we must subscribe to a hidden variable account of quantum mechanics and interpret the wavefunction epistemically, as representative of the possible knowledge an observer can gain about the system. In doing so, we can stipulate through our subsequent quantum model that the state of the system after t_0 is positively correlated to the measurement at t_1 , i.e. that condition (i) holds. Furthermore, we will assume here that we have the power to perform any measurement we like at t_1 , i.e. condition (ii) holds. Let us now consider whether condition (iv) can also hold; that is, whether it is possible to bilk our experiment.

Imagine how a bilking argument against our belief in a retrocausal influence might run. A potential bilker will have to somehow observe the state of the quantum system at some time $t_0 < t < t_1$ and then challenge us to carry out a measurement which is incompatible with this observation. What the bilker will find, however, is a wavefunction description of the system that suggests the system is in an eigenstate of the operator corresponding to the observation that is made at t . Moreover, we know that through this process some probabilistic information about the system will have been lost and thus this wavefunction representation will not be indicative of the state of the system between t_0 and t . Thus the wavefunction description of the state of the system after the bilker's observation will not be what it would have been had the bilker not made that observation; condition (iv) will consequently be violated. Therefore, the very nature of quantum mechanics ensures that any retrocausal effects cannot possibly be bilked of their causes because condition (iv) is perennially violated. Furthermore, we can stipulate further that the intervening observation of the system by the bilker establishes a new correlation between this observation at t and the state of the system between t_0 and t (and with the confidence that this correlation cannot be bilked either).

Thus if we came to the table with a hypothesis about retrocausality in quantum mechanics then we could show that on a certain interpretation of the wavefunction formalism that this particular metaphysical argument against retrocausality does not run. In fact, according to Dummett's analysis of the bilking argument quantum

mechanics has exactly the sort of dynamics we would expect of a retrocausal physical theory; the counterintuitive nature of backwards-in-time causality can hardly be seen as a disadvantage here. We see again that contemporary physics does not preclude a metaphysical picture that allows the possibility of retrocausal influences.

5.4 Keeping up appearances

I hope that it is beginning to become clear the sort of limitations that constrain the form of a picture of reality that allows the possibility of retrocausal influences. We are now in a position to use these constraints, along with the causal and spatiotemporal structures we have taken to be most reasonable, to build a picture of what retrocausality actually involves. At the centre of this discussion will be the role that we play as agents as we interact with, and participate in, the world.

Let us start first and foremost with two conceptions of influence that are commonly conflated when talking about the future: the view that we *change* events and the view that we *affect* events. Consider a claim like the following: by deciding to catch the bus, I changed my day from one in which I was late for work, to one in which I was early. Regardless of one's model of time, there is an inconsistency in thinking that we change events through our actions. For an event to 'change', the event must have been a particular way in the first place. If we were partial to a dynamic view of time in which the future were *unreal*, it would make no sense to think of a future event as being any particular way before it is actual; there is simply no event that is my tardiness which can be changed before I am in fact late. However, we have explicitly signalled our intention to employ the block universe model of time and in such a model we *can* speak of future events as being real and thus it might be possible for an event, one might say, to be a particular way *ab initio*. We might say that my tardiness was an event and that this event changed into my punctuality. But we must be careful here, because if a future event is real, it is in some sense already out there in the four dimensional block. If we *change* it at some point prior to it being a present event for us, we are left with the rather strange question: why was it as it was before we changed it? Why did the four dimensional block contain an event which was my tardiness, which then changed at some point into my punctuality? With respect to the block universe view this question does not make any sense.

Before we confuse matters further, let us take stock and see if we can clarify the above claim. We might do this by saying something like the following: when we say that we change a future event, we mean that we change it from being something that it *could* have been, say my tardiness, to something that it now actually *is*, say my punctuality. Expressing what we mean by change in counterfactual terms lets us sidestep the problems we encountered with the reality of the events under question. However, the notion we have ended up with by doing so has a significant causal ring to it (recall our characterisation of causation in terms of interventions); this is in fact just what we mean when we use the word ‘affect’. I affect my day to be a day in which I am early for work, rather than a day in which I am late. I play a particular role in bringing about the future event and it is wrong to think that I change it from something that it already was. As long as we commit ourselves to the block universe view in which all events in the past, present and future are equally real, then we must think of influence in the ‘affect’ sense. Furthermore, we can now see that this argument is as much relevant to past events as it is relevant to future events. Under no circumstances does it make sense to change the past in any way, since one cannot change something that is already an actual event. Retrocausality is then not about changing the past, rather retrocausality is about affecting the past: playing a role in bringing about a past event.

This analysis is beginning to push us into a position about determinism and the nature of the block universe that may seem highly undesirable; namely, that we have no freedom in choosing our own actions. If we cannot change the future in just the same way that we cannot change the past, and if affectation is merely bringing about an event that in some sense already exists, then it would seem that we are mere spectators of our reality in a rather uninteresting sense. Fortunately, we are not pushed into this position by adopting typically block universe notions as above. Moreover, coming to grips with why this is the case will tie together many of the issues with which we have so far dealt and it will give us our first glimpse at the metaphysical picture of reality that allows for retrocausal influences.

The solution to this seeming incompatibility between the conception of reality as a block universe and our ability as agents to control and manipulate our surroundings lies in thinking of causation as a perspectival notion. According to Price (2007), evidence suggests that causation is indeed a perspectival notion; we have already been introduced to the idea when we were considering the interventionist account

of causation above. The tentative outcome that I flagged of what we called a kind of genealogical account of causation in terms of intervention was that a being who interacted with the world differently to how we interact with the world as agents (i.e. has a different perspective of the world) would have a different concept of causation to the one that we have. Let us consider how we can use this to help us find some sort of compatibility between the block universe view and our causal intuitions.

The essential point to solving this problem is to realise that considering the block universe ‘from the outside’ is availing oneself of a very different perspective of the world to the one which we have while we are inhabiting some spatiotemporal region. The important difference between the two viewpoints is that there is a discrepancy between the parts of the spacetime block that are epistemically accessible from each perspective. The spatiotemporally constrained perspective by which we are bound permits us only limited epistemic accessibility to other spatiotemporal regions. This is significant because it is as spatiotemporally bound agents that we have evolved and it seems reasonable to suggest that we are in possession of a concept of causation that reflects this very fact. Once we imagine ourselves to be omniscient beings that have epistemic access to the whole spatiotemporal block, as we have done in the above analysis of change and affect, it should not come as a surprise that our causal intuitions get confused when we attempt to consider how a spatiotemporally bound agent can deliberate about whether or not to affect a particular event that is already determined from our imagined omniscient perspective. The solution that I am pushing towards here is that it is because we do not *know* which events are determined to occur that we can deliberate, and therefore be causal agents, at all. To reach this conclusion we sequester one final model of the relationship between deliberation and epistemic accessibility, and the role this plays in our concept of causation.

Price (2007, p. 20) sets out “an abstract characterisation of the structural, or functional architecture, of deliberation” with a view to separating out the intrinsic features of deliberation itself from those aspects of deliberation that are a function of our perspective as spatiotemporally bound agents. To begin with, a deliberator must be deliberating over whether to bring about some particular occurrence out of a range of possible occurrences. Following Price, we will call the set of events of which this range consists the *OPTIONS* that the deliberator is considering. The set *OPTIONS* can be thought of as consisting of two subsets: all those occurrences

over which the deliberator has immediate control, the *DIRECT OPTIONS*, and all those occurrences that can be brought about indirectly via the *DIRECT OPTIONS*, the *INDIRECT OPTIONS*. All other events that are not under consideration during the deliberation we will call the *FIXTURES*. An integral subset of the *FIXTURES* is the set of events that the deliberator already knows, or are in principle knowable, at the time of deliberation which we will call the *KNOWABLES*. The *KNOWABLES* must be a subset of the *FIXTURES* since if these events are knowable to the deliberator at the time of deliberation, then they cannot be under consideration to be brought about and thus cannot be part of the set *OPTIONS*. For this reason, all the events in *OPTIONS* must fall into the set we will call *UNKNOWABLES*. Thus a deliberator makes two dichotomous distinctions: the distinction between *FIXTURES* and *OPTIONS*; and the distinction between *KNOWABLES* and *UNKNOWABLES*. The set *KNOWABLES* is a subset of *FIXTURES* and the set *OPTIONS* is a subset of *UNKNOWABLES*. Let us now consider how spatiotemporally bound deliberators such as ourselves might map these distinctions onto the past and the future.

Considering the future first, we are going to want to say that much of the future belongs to the set *FIXTURES*. This is largely due to the finite nature of deliberation: since we do not deliberate about bringing about the whole future all at once, there are then many future occurrences that we take as part of the fixed background during the deliberative process. It also seems as given that the set *DIRECT OPTIONS* must also be comprised of future events. We can attribute this to the fact that we are temporally constrained agents of a particular sort; the set *DIRECT OPTIONS* consists of our immediate actions and we simply cannot deliberate about whether to bring about our past actions, only our future actions. Further to this, we might want to say that the set *INDIRECT OPTIONS* also is comprised exclusively of future events, but this would be so only if we were committed to classifying all past events as belonging to the set *FIXTURES*. Ordinarily, this is exactly how we consider past events: as fixed. This is for the most part a function of the fact that we consider the past as knowable in principle, and as we have seen above, the set *KNOWABLES* is a subset of the set *FIXTURES*. But is it the case that our spatiotemporally bound perspective commits us to the past being fixed?

If such a commitment is indeed a function of the fact that we consider the past as knowable in principle, then it would seem that the possibility of the past being unknowable in principle would purge us of this commitment. Recall that

this is exactly the condition we found to be suitable to avoid the bilking argument in the above analysis of Dummett: an agent is immune to having a belief in a particular retrocausal correlation bilked if the past effect in question is epistemically inaccessible to the agent at the time of the causal action. In the language of our current analysis, if some past event belongs to the set UNKNOWABLES then it does not *necessarily* belong to the set FIXTURES, and an agent may then believe it to belong to the set INDIRECT OPTIONS. As we noted above, the very nature of quantum mechanics ensures that it *is* immune to the bilking argument. Thus, in the right circumstances, there is information about the past of some quantum systems that is epistemically inaccessible *in principle*! If this is the case then it is a live possibility that the set INDIRECT OPTIONS contains some events which are past; or rather, the architecture of deliberation does not rule out the possibility of bringing about the past on analytic grounds.

This schematic of where retrocausality fits in to the structure of deliberation highlights an important feature of a metaphysical picture that allows retrocausal influences: that agents within such a reality will always deliberate towards the future, i.e. the set DIRECT OPTIONS will always be comprised of future events. Thus retrocausality is not deliberation towards the past, or in other words, it is not our normally directed causation in the reverse temporal direction.

The way that any particular agent divides the set of all events into FIXTURES and OPTIONS, KNOWABLE and UNKNOWABLE and past and future will depend completely upon the agents spatiotemporal perspective. For spatiotemporally constrained agents such as ourselves, there is a specific recipe for how these distinctions are made which is a function of the way we have evolved from within the spacetime block. If we imagine ourselves as omniscient beings who are observing the events in the spacetime block ‘from the outside’, there will be no past or future (though there may be past and future directions along the temporal axis of the block) and all the events will be in the set KNOWABLE and thus in the set FIXTURES. This is how we can imagine the spacetime block to be entirely determined without having this intuition be in conflict with our usual sense of free choice in the deliberative process; these are vastly different perspectives and causality is perspectival. It is the extent of our ignorance, of both the future and of the complete set of prior causes of our actions, that creates the illusion, so to speak, of free choice. This is where we then strike a harmony between our causal intuitions, such as deliberation, and the

intuition that future events are fixed within a deterministic framework. The crucial element is to realise that we, as spatiotemporally bound agents, are constrained in our epistemic access to the events in spacetime. We will meet this picture again explicitly in §6.8.1 in the context of retrocausal quantum mechanics.

5.5 A retrocausal picture

This then is the package of metaphysical ideas that combine to give a picture that is consistent with the possibility of retrocausality. We begin with two uncontroversial metaphysical foundations in the block universe model of time and the interventionist account of causation. We then remove two potential obstacles originating in our ordinary temporal intuitions: we realise that we have no evidence to suggest our macroscopic asymmetric causal intuitions can be extrapolated to the microscopic realm and we realise that we do not necessarily have epistemic access to the past independent of our own future actions. With these obstacles gone, the emerging picture of a temporally and causally symmetric reality viewed from an epistemically limited vantage point concords well with the possibility of retrocausality. A significant aspect of this assembly of ideas is that none of the included elements are precluded by current physical theory. Indeed, if anything, these elements are supported by the structure of at least one of our best physical theories: quantum mechanics.

Before we move on, however, let us recall the sentiment of the Maudlin quote with which we began this chapter. While Maudlin is clearly correct in noticing that retrocausality is fundamentally at odds with the metaphysical picture of the past generating the future, this by no means renders retrocausality metaphysically untenable. Given the right mix of some reasonable metaphysical and epistemological ingredients, an alternative picture of reality arises that is consistent with the possibility of retrocausality. Moreover, the economical cost of these ingredients cannot outweigh the interpretational problems associated with the rejection of local hidden variables, simply for the fact that we were given all these ingredients for free by the metaphysical structure of our existing physical theories and the epistemological structure of our experiences. We turn now to the task of defending retrocausality against Maudlin's challenge.

Causal Symmetry and the Transactional Interpretation of Quantum Mechanics

So far we have examined retrocausality as a solution to the interpretational problems for quantum mechanics raised by Bell's theorem and as a metaphysical possibility within contemporary physics. However, one of the most significant obstacles for retrocausal approaches to quantum mechanics is the objection levelled at John Cramer's (1986) transactional interpretation of quantum mechanics by Tim Maudlin (2002), who claims that his objection poses a problem for "any theory in which both backwards and forwards influences conspire to shape events". This chapter is an examination of Maudlin's objection to retrocausality.

6.1 Introduction

The examination proceeds as follows. I begin in §6.2 with an introduction to Wheeler and Feynman's (1945) attempted time symmetric formulation of classical electrodynamics, from which the transactional interpretation of quantum mechanics originates. I then introduce in §6.3 Cramer's extension of the Wheeler-Feynman formalism to a retrocausal transaction mechanism for modelling quantum processes. §6.4 sets out the details of the transactional interpretation and I briefly mention there some of the advantages Cramer's theory has over the Copenhagen interpretation of quantum mechanics: most notably that the retrocausal structure allows a 'zigzag' causal explanation of the nonlocality associated with Bell-type quantum systems. In §6.5 I set out the details of Maudlin's inventive thought experiment that constitutes his objection to Cramer's theory. I examine in §6.6 some replies that have been made in response to Maudlin's objection defending the transactional interpretation.

In §6.7 I offer my own analysis of Maudlin’s experiment according to the transactional interpretation with a view to showing that, despite the putative defences considered, there is still a problem to be overcome. What is lacking in Cramer’s theory is a causal structure that can constrain uniquely the behaviour of a quantum system and this is exactly the problem that Maudlin’s experiment emphasises. I diverge from Maudlin, however, in the justification for why the transactional interpretation suffers this shortcoming. I claim in §6.8 that it is the failure of the transactional interpretation to ensure causal symmetry that is impeding such unique determination of behaviour. In contrast, Maudlin attributes this shortcoming to retrocausality itself and emphasises an apparently fundamental incongruence between retrocausality and his own “metaphysical picture of the past generating the future”. I present an argument that it is Maudlin’s assumption about the appropriateness of this metaphysical picture that is problematic here, and not retrocausality.

6.2 The Wheeler-Feynman absorber theory of radiation

Our narrative begins with a problem of classical electrodynamics: an accelerating electron emits electromagnetic radiation, and through this process the acceleration of the electron is damped. Various attempts were initially made to account for this phenomenon in terms of the classical theory of electrodynamics but largely these lacked either empirical adequacy or a coherent physical interpretation. Wheeler and Feynman (1945) set out to remedy this situation by reinterpreting Dirac’s (1938) theory of radiating electrons. I will make no attempt here to give an analysis of this problem, nor of its ensuing evolution. What is important for our purposes is the nature of the interpretation that Wheeler and Feynman proffer as a resolution, for it is this interpretation that is the motivation for the transactional interpretation of quantum mechanics.

The core of Wheeler and Feynman’s absorber theory of radiation is a suggestion that the process of electromagnetic radiation should be thought of as an interaction between a source and an absorber rather than as an independent elementary process.¹ Wheeler and Feynman imagine an accelerated point charge located within an

¹Such an idea was suggested, for instance, by Tetrode (1922) and also by Lewis (1926):

[A]n atom never emits light except to another atom, and... it is as absurd to think of light emitted by one atom regardless of the existence of a receiving atom as it

absorbing system and consider the nature of the electromagnetic field associated with the acceleration. An electromagnetic disturbance initially travels outwards from the source and perturbs each particle of the absorber. The particles of the absorber then generate together a subsequent field. According to the Wheeler-Feynman view, this new field is comprised of half the sum of the retarded (forwards-in-time) and advanced (backwards-in-time) solutions to Maxwell's equations. The sum of the advanced effects of all the particles of the absorber then yields an advanced incoming field that is present at the source simultaneous with the moment of emission. The claim is that this advanced field exerts a finite force on the source which has exactly the required magnitude and direction to account for the observed energy transferred from source to absorber; this is Dirac's radiative damping field. In addition, when this advanced field is combined with the equivalent half-retarded, half-advanced field of the source, the total observed disturbance is the full retarded field known from experience to be emitted by accelerated point charges.

The crucial point to note about the Wheeler-Feynman scheme is that due to the *advanced* field of the absorber, the radiative damping field is present at the source at exactly the time of the initial acceleration. Quite simply, if a retarded electromagnetic disturbance propagates for a time t before meeting the absorber then the absorber will be a distance ct from the source. The advanced field propagates with the same speed c across the same distance and thus will arrive at the source exactly time t *before* the absorber field is generated, i.e. at the time of the initial acceleration. If we think of this four dimensionally (in block universe terms) it is clear to see that the advanced field does not simply propagate the same distance as the source field, it propagates across the very same spacetime points as the initial disturbance.² It is this idea of time symmetric radiation that is at the core of the transactional interpretation of quantum mechanics.

would be to think of an atom absorbing light without the existence of light to be absorbed. I propose to eliminate the idea of mere emission of light and substitute the idea of transmission, or a process of exchange of energy between two definite atoms or molecules. (Lewis, 1926, p. 24)

²According to Price (1991), the fact that the retarded and advanced waves cross the same spacetime indicates that they are in fact one and the same electromagnetic disturbance.

6.3 The quantum handshake

Cramer's (1986) transactional interpretation is a retrocausal model of quantum mechanics that extends the Wheeler-Feynman formalism beyond electrodynamics. Cramer suggests that the description of the emission and absorption of electromagnetic radiation in the Wheeler-Feynman scheme can be adopted to describe the microscopic exchange of a single quantum of energy, momentum, *etc.*, between and within quantum systems. This time symmetric interpretation of the quantum mechanical formalism not only provides an action-by-contact explanation of the non-locality found in the EPRB experiment but also constitutes an attempt to alleviate some of the interpretational problems of the Copenhagen interpretation in general. Before we address how this is achieved, let us consider the transaction mechanism at the core of the transactional interpretation.

Imagine a quantum emitter such as a vibrating electron or atom in an excited state. According to Cramer, when a single quantum is to be emitted (a photon, in these cases) the source produces a radiative field. Analogously to the Wheeler-Feynman description, this field propagates outwards in all directions of four dimensional spacetime, i.e. in all three spatial dimensions and both forwards (retarded field) and backwards (advanced field) in the temporal dimension. When this field encounters an absorber, a new field is generated that likewise propagates in all directions of four dimensional spacetime. The retarded field produced by the absorber exactly cancels the incident retarded field produced by the emitter for all times after the absorption of the photon. The advanced field produced by the absorber propagates backwards in time across the same spacetime interval as the incident wave to be present at the emitter at the instant of emission. The advanced field produced by the absorber exactly cancels the advanced field produced by the emitter and thus there is neither a net field present after the time of absorption nor before the initial emission; only between the emitter and the absorber is there a radiative field.

Cramer describes the field that travels from the source to the absorber as an "offer" wave and the field that returns from the absorber to the emitter as a "confirmation" wave. The transaction is completed with a "handshake": the offer and confirmation waves combine to form a four dimensional standing wave between emitter and absorber. The conditions at the emitter and absorber at the time of emission and absorption respectively are the boundary conditions that determine whether or

not a transaction can take place and, if so, the probability of that transaction occurring. The amplitude of the confirmation wave which is produced by the absorber is proportional to the local amplitude of the incident wave that stimulated it and this, in turn, is dependent on the attenuation it received as it propagated from the source. It is the boundary conditions at both ends of the transaction that define when a transaction can be completed. A cycle of offer and confirmation waves “repeats until the response of the emitter and absorber is sufficient to satisfy all of the quantum boundary conditions... at which point the transaction is completed” (1986, p. 662). Many confirmation waves from potential absorbers may converge on the emitter at the time of emission but the quantum boundary conditions can usually only permit a single transaction to form. Any observer who witnesses this process would perceive only the completed transaction, which would be interpreted as the passage of a particle (e.g. a photon) between emitter and absorber.

There are in fact two complementary descriptions of the transaction process lurking side by side here: on the one hand there is a description of the physical process, consisting of the passage of a particle between emitter and absorber, that a temporally bound experimenter would observe; and on the other hand there is a description of a dynamical process of offer and confirmation waves that is instrumental in establishing the transaction. This latter process clearly cannot occur in an ordinary time sequence, not least because our temporally bound observer by construction cannot detect any offer or confirmation waves. Cramer suggests that the ‘dynamical process’ be understood as occurring in a “pseudotime” sequence:

The account of an emitter-absorber transaction presented here employs the semantic device of describing a process extending across a lightlike or a time-like interval of space-time as if it occurred in a time sequence external to the process. The reader is reminded that this is only a pedagogical convention for the purposes of description. The process is atemporal and the only observables come from the superposition of all “steps” to form the final transaction. (1986, p. 661, fn. 14)

These steps are of course the cyclically repeated exchange of offer and confirmation waves which continue “until the net exchange of energy and other conserved quantities satisfies the quantum boundary conditions of the system” (Cramer, 1986, p. 662). There is a strong sense here that any process described as occurring in pseudotime is *not a process at all* but, as Cramer reminds, merely a “pedagogical-

cal convention for the purposes of description”. The role that pseudotime plays in Cramer’s theory will be of major concern for us in this analysis and we will see in §6.7 that the ontological status of Cramer’s posited pseudotemporal sequence is far from transparent. For now, however, let us consider how this transaction mechanism underpins Cramer’s transactional interpretation of quantum mechanics.

6.4 The transactional interpretation

Cramer utilises the principled framework of the Copenhagen interpretation to characterise his transactional interpretation. Recall that the Copenhagen interpretation can be characterised in terms of a clutch of core principles, including Heisenberg’s indeterminacy relation, the Born rule, Bohr’s principle of complementarity and the epistemic reading of the wavefunction. The purpose of these principled elements is to provide a physical picture of quantum systems given the formalism of quantum mechanics; Cramer likewise constructs the transactional interpretation from principles to serve this end.

To begin with, the statistical interpretation of the formalism embodied in the Born rule remains unchanged from the Copenhagen interpretation. This is a consequence of the fact that during the transaction process the confirmation wave traverses the very same spacetime as the offer wave, only in reverse: the amplitude of the advanced component of the confirmation wave arriving back at the emitter is proportional to the time reverse (or complex conjugate) of the amplitude of the initial offer wave evaluated at the absorber. Thus, the total amplitude of the confirmation wave is just the absolute square of the initial offer wave (evaluated at the absorber), which yields the Born rule. Since the Born rule arises as a product of the transaction mechanism, there is no special significance attached to the role of the observer in the act of measurement. The ‘collapse of the wave function’ is interpreted as the completion of the transaction. Thus both the indeterminacy relation and the principle of complementarity are no longer fundamentally related to the process of observation but rather dissolve into a single feature of the transaction mechanism: in satisfying the boundary conditions, the transaction can project out and localise only one of a pair of conjugate variables.

According to Cramer, the biggest bifurcation between the Copenhagen and transactional interpretations is centred around the physical significance of the wavefunc-

tion. As a function of the principle of complementarity, the completeness of the quantum formalism and the need to avert worries about the nonlocality of the collapse process, the wavefunction according to the Copenhagen interpretation can be thought of as simply “a mathematical description of the state of observer knowledge” (Cramer, 1988, p. 228).³ In contrast, the transactional interpretation takes the wavefunction to be a real physical wave with spatial extent.⁴ The wavefunction of the quantum mechanical formalism is identical with the initial offer wave of the transaction mechanism and the collapsed wavefunction is identical with the completed transaction. Quantum particles are thus not to be thought of as represented by the wavefunction but rather by the completed transaction, of which the wavefunction is only the initial phase. As Cramer explains:

The transaction may involve a single emitter and absorber or multiple emitters and absorbers, but it is only complete when appropriate boundary conditions are satisfied at all loci of emission and absorption. Particles transferred have no separate identity independent from the satisfaction of these boundary conditions. (1986, p. 666)

Though there is much formal overlap between particular elements of the Copenhagen and transactional interpretations, Cramer points out that giving objective reality to the wavefunction “colors all the other elements of the interpretation” leading to a vastly different physical picture of the quantum world. Let us consider this physical picture with a concrete example.

Consider a radioactive source, S, sitting between two absorbers, A and B, constrained to emit a single β -particle either to the left or to the right (Figure 6.1). According to the transactional interpretation, the process of β -particle emission can be described in terms of offer and confirmation waves, the initial offer wave being the wavefunction of the quantum mechanical formalism. The wavefunction “is a real

³Cramer’s reading of the Copenhagen interpretation in this respect is somewhat contentious. The “knowledge interpretation” of the wavefunction is claimed to be an integral element of the Copenhagen interpretation by Heisenberg (1955) but this may be in conflict with the way Bohr envisaged the wavefunction. See Howard (2004) for an excellent discussion of this issue.

⁴Recent work by Kastner (2010) and Kastner and Cramer (2010) suggests a potentially improved reading of the transactional interpretation where the wavefunction is considered as “residing in a ‘higher’ or external ontological realm corresponding to the Hilbert space of all quantum systems involved”. This is an interesting and potentially fruitful avenue for avoiding the problems I outline below concerning Cramer’s realistic interpretation. I unfortunately do not take account of this improved reading here.

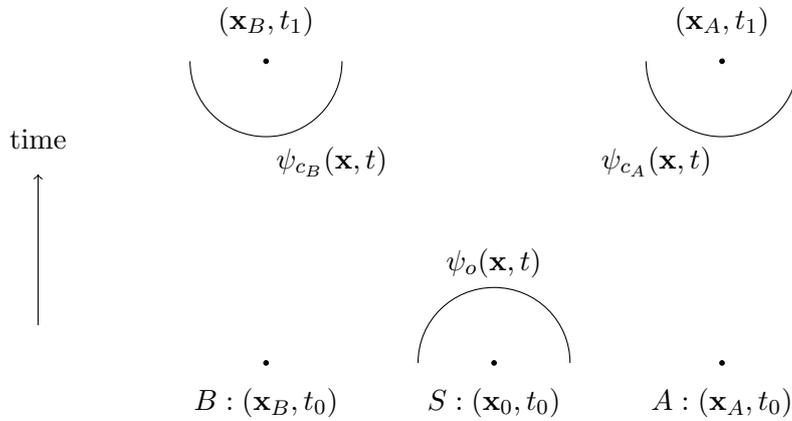


Figure 6.1: Offer and confirmation waves in the transactional interpretation

physical wave generated by the emitter, and travels through space to the final absorber as well as to many other spacetime loci and many other potential absorbers” (1986, p. 667). Thus at the time of emission, t_0 , an offer wave is produced which propagates towards each absorber as well as forwards and backwards in time. Upon being stimulated by this offer wave at time t_1 , the absorbers A and B each produce a confirmation wave that propagates backwards in time (among other directions) to the radioactive source; the amplitude of each confirmation wave evaluated at the source is proportional to the modulus squared of the amplitude of the offer wave evaluated at the respective absorber (i.e. $|\psi_{c_i}(\mathbf{x}_0, t_0)| \propto |\psi_o(\mathbf{x}_i, t_1)|^2$ for $i = A, B$). These confirmation waves provide the emitter, so to speak, with a Born probability measure over which the likelihood of each particular transaction occurring can be quantified. In the same way that the absorber responds to the initial offer wave, the emitter responds to this subsequent confirmation wave and this cycle continues.

Let us say that the transaction is completed between the radioactive source and absorber A, i.e. a four dimensional standing wave emerges between S at t_0 and A at t_1 . The components of the wavefunction which permeate the spatiotemporal regions that are not between S at t_0 and A at t_1 do not “disappear”. Rather these components “are only virtual in the sense that they transfer no energy or momentum and participate in no transaction”. Moreover, “the emergence of this transaction does not occur at any particular location in space or at some particular instant in time, but rather forms along the entire four-vector that connects the emission locus

with the absorption locus” (1986, p. 667). This four dimensional standing wave is then interpreted as the emission of a β -particle to the left by S at t_0 and the subsequent absorption of this β -particle by A at t_1 .

The transactional interpretation of the quantum formalism allows the resolution of some of the most worrying aspects of the Copenhagen interpretation. Since we do not require the disappearance of the initial wavefunction upon completion of the transaction, the transactional interpretation alleviates the need to resort to an epistemic interpretation of the wavefunction, which Cramer (1988, p. 228) finds “intellectually unappealing”, to account for the nonlocality associated with wavefunction collapse.⁵ In addition, the transactional interpretation subverts the dilemma at the core of the EPR argument (§4.5) by permitting the simultaneous reality of incompatible operators: the wavefunction, according to the transactional interpretation,

brings to each potential absorber the full range of possible outcomes, and all have “simultaneous reality” in the EPR sense. The absorber interacts so as to cause one of these outcomes to emerge in the transaction, so that the collapsed [wavefunction] manifests only one of these outcomes. (1986, p. 668).

Most importantly, however, the transactional interpretation employs both retarded and advanced waves, and in doing so admits the possibility of providing a ‘zigzag’ explanation of the nonlocality associated with the EPRB experiment. The boundary conditions that influence the formation of a completed transaction include both those at the emitter as well as the future absorbers. It is this feature that makes the transactional interpretation a retrocausal model of quantum mechanics. Moreover, it is this feature that enables the combination of two or more local influences to yield a nonlocal influence, which allows an action-by-contact description of the EPRB experiment.

While it at least appears as though the transactional interpretation goes some way to resolving the interpretational issues of the Copenhagen interpretation, it is in fact not without its own points of contention.

⁵See fn. 3.

6.5 Maudlin's objection

Maudlin (2002) outlines a selection of problems that arise in Cramer's theory as a result of the pseudotemporal account of the transaction mechanism: processes important to the completion of a transaction take place in pseudotime only (rather than in real time) and thus cannot be said to have taken place at all. Since a temporally bound observer can only ever perceive a completed transaction, i.e. a collapsed wavefunction, the uncollapsed wavefunction never actually exists. Since the initial offer wave is identical to the wavefunction of the quantum formalism, any ensuing exchange of advanced and retarded waves required to provide the quantum mechanical probabilities, according to Maudlin, also do not exist. Moreover, Cramer's exposition of the transaction mechanism seems to suggest that the stimulation of sequential offer and confirmation waves occurs deterministically, leaving a gaping hole in any explanation the transactional interpretation might provide of the stochastic nature of quantum mechanics. Although these problems are significant, Maudlin admits that they may indeed be peculiar to Cramer's theory. Having said this, Maudlin also sets out a more general objection to retrocausal models of quantum mechanics which he claims to pose a problem for "any theory in which both backwards and forwards influences conspire to shape events" (2002, p. 201).

Maudlin's main objection to the transactional interpretation hinges upon the fact that the transaction process depends crucially on the fixity of the absorbers "just sitting out there in the future, waiting to absorb" (2002, p. 199); one cannot presume that present events are unable to influence the future disposition of the absorbers. Let us consider Maudlin's own thought experiment designed to illustrate this objection. Consider again our radioactive source constrained to emit a β -particle either to the left or to the right. To the right sits absorber A at a distance of 1 unit. Absorber B is also located to the right but at a distance of 2 units and is built on pivots so that it can be swung around to the left on command (Figure 6.2(i)). A β -particle emitted at time t_0 to the right will be absorbed by absorber A at time t_1 . If after time t_1 the β -particle is not detected at absorber A, absorber B is quickly swung around to the left to detect the β -particle after time $2t_1$ (Figure 6.2(ii)).

According to the transactional interpretation, since there are two possible outcomes (detection at absorber A or detection at absorber B), there will be two confirmation waves sent back from the future, one for each absorber. Furthermore, since

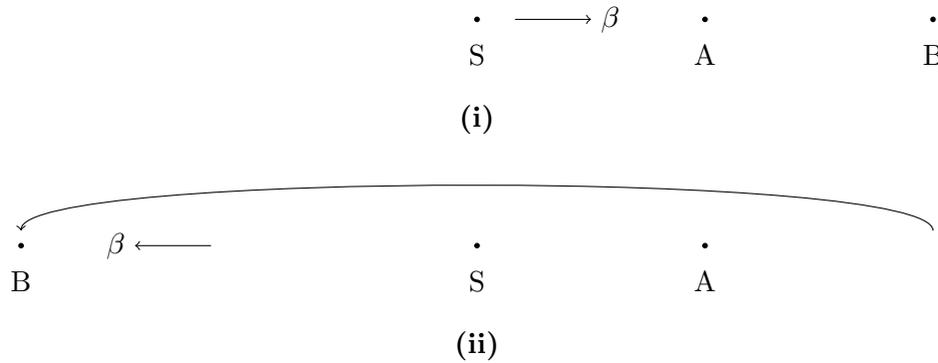


Figure 6.2: Maudlin's thought experiment

it is equally probable that the β -particle be detected at either absorber, the amplitudes of these confirmation waves should be equal. However, a confirmation wave from absorber B can only be sent back to the emitter if absorber B is located on the left. For this to be the case, absorber A must not have detected the β -particle and thus the outcome of the experiment must already have been decided. The incidence of a confirmation wave from absorber B at the emitter *ensures* that the β -particle is to be sent to the left, even though the amplitude of this wave implies a probability of a half of this being the case. As Maudlin states so succinctly, “Cramer’s theory collapses”.

It is clear to see that this challenge to retrocausality must be considered seriously if a proposed retrocausal mechanism is to be successful. The key challenge from Maudlin is that any retrocausal mechanism must ensure that the future behaviour of the system transpires consistently with the spatiotemporal structure dictated by any potential future causes: “stochastic outcomes at a particular point in time may influence the future, but that future itself is supposed to play a role in producing the outcomes” (2002, p. 197). In the transactional interpretation the existence of the confirmation wave itself presupposes some determined future state of the system with retrocausal influence. However, with standard (i.e. forwards-in-time) stochastic causal influences affecting the future from the present, a determined future *may not necessarily be guaranteed* in every such case, as shown by Maudlin’s experiment. Before we go on to examine this objection in more detail, let us first consider some responses that have been put forward in defence of Cramer’s theory.

6.6 Cramer defended

I wish to examine here three specific defences of the transactional interpretation due to Berkovitz (2002), Kastner (2006) and Marchildon (2006). A review of these defences will not only provide a good exercise in exploring the details of the transactional interpretation, but will assist us in getting to the source of the issues highlighted by Maudlin's challenge.

Maudlin's objection has been formulated by Berkovitz (2002) in terms of the varying conceptualisations of the probabilities involved in the experiment. More specifically, Berkovitz believes that the deviation between the long-run frequencies of measurement outcomes and their objective probabilities is at the core of the objection. Berkovitz defends the transactional interpretation by showing that causal loops of the type found in Maudlin's experiment need not obey the assumptions about probabilities that are common in linear causal situations. To illustrate this claim about causal loops, Berkovitz considers a simple coin toss.

Let the event P be the tossing of a fair coin, and let this be an indeterministic cause of event Q, the coin landing 'heads'. Let event R be the perception of the coin landing 'heads' deterministically caused by event Q. Since the coin is fair, the long-run frequency of event Q with respect to event P is $\frac{1}{2}$. However, if one considers the long-run frequency of event Q with respect to both P *and* R, then this frequency is 1; every time event P occurs with event R, Q must have occurred. The probability of event Q with respect to P and R is called by Berkovitz a biased probability. Berkovitz argues that within causal loops of the type found in Maudlin's experiment the probabilities are always biased. Thus one should not expect the long-run frequencies to correspond with any unbiased probabilities; there is no inconsistency in a deviation between these quantities.

This example can be translated in a straightforward manner to the language of Maudlin's experiment. Event P is the radioactive β -decay, event Q (indeterministically caused by P) is the emission of the β -particle to the left and event R (deterministically caused by Q) is the detection of the β -particle on the left. Recall that an integral element of Maudlin's objection is that the existence of the confirmation wave on the left *ensures* event Q, but the information contained within the confirmation wave itself suggests event Q has a probability of only $\frac{1}{2}$. With respect to only event P, event Q has a long-run frequency of $\frac{1}{2}$, but with respect to both

P and R this biased probability is 1. It is not inconsistent for these quantities to deviate, therefore Berkovitz claims Cramer's theory is not inconsistent.

Berkovitz does not consider Cramer's pseudotemporal account of the transaction mechanism significant, preferring to think of the cycle of offer and confirmation waves in terms of causal connections which are part of a four dimensional block universe. While Berkovitz has claimed to show the legitimacy of the causal loop in Maudlin's experiment, by overlooking Cramer's pseudotemporal account of the transaction mechanism Berkovitz has neglected to address exactly how the pseudotemporal account *can* be consistent with the four dimensional block universe. Berkovitz returns to the transactional interpretation in his (2008) where he recognises that the pseudotemporal account of the transaction mechanism jeopardises the explanatory value of the theory. However, the ontological nature of the transaction mechanism is once again left to one side in his analysis.

Kastner (2006) has expanded on Berkovitz' approach with a view to eliminating pseudotime from the transactional interpretation. Kastner begins by noting that in the transactional interpretation a *complete* set of absorbers is not necessary; it is possible for no confirmation wave to be received from the left of the radioactive source in Maudlin's experiment. Kastner differentiates between the initial states of the radioactive source in the two situations where (i) a confirmation wave is received from both the right and the left absorbers (absorbers A and B respectively), and (ii) a confirmation wave is received from only the right absorber (absorber A only). It is clear that if a confirmation wave is received from both the left and right then it is the case that the β -particle will be emitted to the left. Recall that Maudlin claims this to be inconsistent with the information contained in the confirmation wave from the left.

In a similar fashion to the analysis of Berkovitz, Kastner emphasises the disparity of probabilities as the heart of Maudlin's objection to the transactional interpretation. However, given the initial state, according to Kastner, this disparity can be explained. The probability of emission to the left in the case where a confirmation wave is received from both the left and the right is $\frac{1}{2}$ according to the information contained in each confirmation wave. However, the probability of this being the initial state of the emitter is also $\frac{1}{2}$ since there are two equally probable initial states. Thus, using the standard probabilistic expression, the probability of emission to the

left given the initial state is

$$\frac{P(L \& \psi)}{P(\psi)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1,$$

where L is emission to the left and ψ is the initial state.

These two initial states of Maudlin's experiment can be imagined, according to Kastner, as belonging to two distinct worlds, which share only the offer and confirmation waves between the emitter and absorber A in common. Kastner proposes that the incipient transaction corresponding to the offer and confirmation waves between the emitter and absorber A can be thought of as an unstable bifurcation line between the two worlds. The success or failure of this transaction determines which world the system "enters". Suppose the incipient transaction between the emitter and absorber A fails. If this is the case, absorber A does not detect the β -particle and absorber B is swung around to the left where it is now able to emit a confirmation wave. What would otherwise have been a null outcome becomes a realised transaction.

Kastner points out that this account must "abandon the idea that there is cyclic 'echoing' between absorber B and the emitter if such echoing is taken as reflective of an uncertainty in outcome" (2006, p. 14). The failure of the bifurcating transaction, i.e. that between the emitter and absorber A, makes the outcome of emission to the left certain. Moreover, the information contained in the confirmation wave received from absorber B indicates a probability of a half of this being the case. This is not inconsistent due to the above analysis and, in fact, shows that each confirmation wave reflects the probability structure across both possible worlds, demonstrating the holistic structure of quantum mechanics.

Marchildon (2006) proposes another defence of the transactional interpretation against Maudlin's objection. He begins by supposing another absorber, say C, is situated on the left of the radioactive source in Maudlin's thought experiment at a distance larger than that of absorber B from the source. If this is the case, then the emitter will receive a confirmation wave from absorber C on the left and Maudlin's experiment will proceed as usual. Marchildon then proposes removing absorber C and considering the absorption properties of the long distance boundary conditions. If it is postulated that the universe is a perfect absorber of all radiation then the

presence of absorber C is irrelevant; a confirmation wave from the left will always be received by the radioactive source at the time of emission and it will encode the correct probabilistic information. This enables the transactional interpretation to remain consistent in Maudlin's experiment. On the assumption that the universe is a perfect absorber, the transactional interpretation correctly predicts that the β -particle will be emitted to the left half of the time. It remains the case, however, that the transaction is completed with absorber B only if it is situated on the left. According to Marchildon, "although the confirmation wave coming from the left originates from the remote absorber just as often as it originates from B, the transaction is never completed with the remote absorber" (2006, p. 12).

Although there is a varied focus to each of these defences of Cramer, it is clear that the problematic element of the transactional interpretation is the causal structure of the pseudotemporal account of the transaction mechanism. In the next section I offer an analysis of Maudlin's experiment according to this pseudotemporal account from the perspective of the block universe model. In doing so I hope to show why Maudlin's experiment still poses a problem for Cramer's theory despite these defences. The underlying problem is that Cramer's theory fails to provide a sufficient causal structure to constrain uniquely the behaviour of the system. While I think Maudlin has successfully isolated this shortcoming, in §6.8 I challenge his justification for why this is the case in the transactional interpretation.

6.7 Maudlin's experiment in four dimensions

The central claim with which Berkovitz and Kastner are concerned is the disparity between the probability of emission to the left as determined by the amplitude of the confirmation wave from the left absorber and the expected probability given that a confirmation wave arrives from the left. Above I characterised Maudlin's objection in a different manner: Maudlin's key challenge is that any retrocausal mechanism must ensure that the future behaviour of the system transpires consistently with the spatiotemporal structure dictated by any potential future causes. An instructive way to analyse the causal structure of Maudlin's experiment is four dimensionally.

Consider once again Maudlin's experimental setup and let us imagine a β -particle emission to the right (i.e. towards absorber A) according to the transactional interpretation. Figure 6.3 represents that part of the transaction process that occurs in

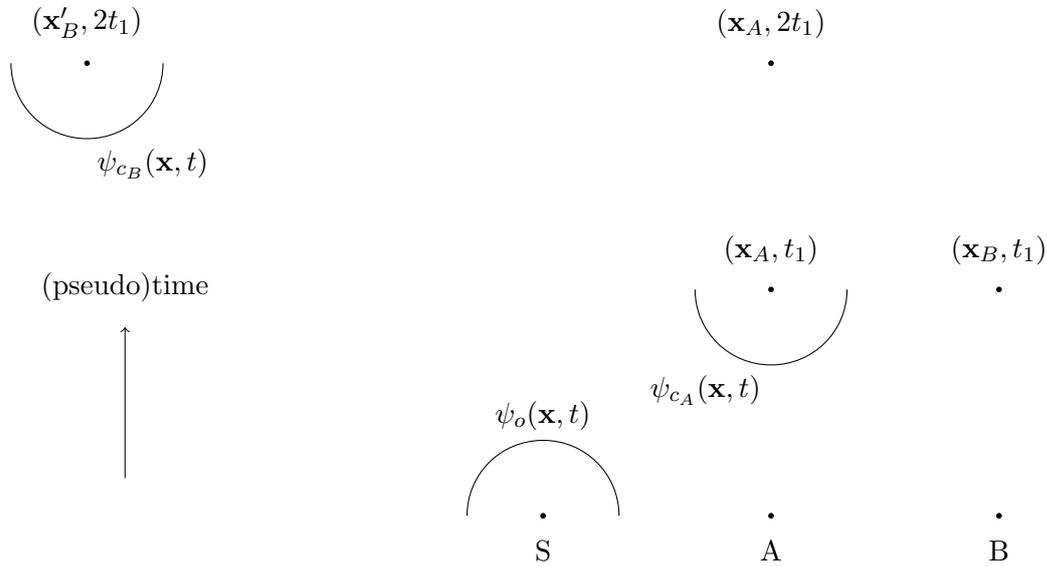


Figure 6.3: Maudlin's thought experiment according to the transactional interpretation

pseudotime (a sort of 'space-pseudotime' diagram). An offer wave is emitted from the radioactive source at time t_0 . If we initially ignore the conditional nature of the event structure of the experiment, we can imagine this offer wave stimulating two confirmation waves from absorbers A and B, each confirmation wave originating from the respective potential absorber positions in spacetime. The particular transaction process we are considering determines that a four dimensional standing wave emerge between absorber A at time t_1 and the source S at time t_0 , which is interpreted as the emission of a β -particle at S and the absorption of this particle at A. The passage of the β -particle, whose four-vector emerges atemporally over the entire locus of the transaction, is a process of spacetime while the transaction mechanism itself is a process of pseudotime. If we now consider the conditional nature of the event structure in time, due to the absorption event at absorber A, absorber B will remain on the right. Curiously the four dimensional 'space-pseudotime' block contains an event structure (i.e. absorber B swinging to the left) which the four dimensional spacetime block does not. Both absorber A and absorber B vie *in pseudotime* to participate in the completed transaction but, once the transaction emerges for one absorber only, the standing wave that is formed is a standing wave *in spacetime*. We should acknowledge at this point that Cramer attempts to allevi-

ate such a worry by suggesting pseudotime to be “a pedagogical convention for the purposes of description” (see §6.3).

However, this worry is nonetheless compounded by Cramer's insistence that the initial offer wave, identical to the wavefunction of the quantum formalism, is a real wave that propagates through space. Recall that the components of this offer wave that do not participate in any eventual transaction are described by Cramer as “virtual” in that they transfer no energy or momentum. What Cramer fails to account for is the fact that these virtual components do contribute something quite important to the transaction mechanism: a putative causal structure. The role that is played by those components of the wavefunction that are not emitted in the direction of the eventual absorber is to stimulate virtual confirmation waves which in turn provide the emitter with the relevant Born probability measure over all future possibilities. This potential causal influence, however, originates from a ‘space-pseudotime’ event structure that is not necessarily representative of the event structure in the future of the quantum system; the probability measure is constrained by objects that may not physically be there! There is then something very strange in claiming that the virtual cycles of offer and confirmation waves play a causal role in constraining the event structure of spacetime. There seems to be a mismatch between the causal structure dictated by the initial conditions and the causal structure dictated by the actual evolution of the system which turns on the obscure ontological status of the pseudotemporal process.

Maudlin's objection is a potent one; the pseudotemporal account of the transaction mechanism in the transactional interpretation resists straightforward clarification. This somewhat complicates the defences examined in the last section, particularly those of Berkovitz and Marchildon. Berkovitz does consider the cycle of offer and confirmation waves between emitters and absorbers in the context of a four dimensional block universe (2002, p. 242), but does so without considering the reality of these entities within the spacetime block. By playing an important role in the transaction mechanism, the offer and confirmation waves have causal significance in Berkovitz' causal loops. However, it is this causal significance that is called into question by Maudlin's objection rather than the consistency of the causal loops.

In contrast, Marchildon eschews any causal significance of the pseudotemporal offer and confirmation waves by assuming the universe to be a perfect absorber; there

will always be *some* confirmation wave returning to the source at the time of emission from every direction, which can play the role of providing the Born probability measure. However, this does not do justice to the pseudotemporal account of the transaction mechanism. Recall that “the emergence of [the completed] transaction does not occur at any particular location in space or at some particular instant in time, but rather forms along the entire four-vector that connects the emission locus with the absorption locus”. Thus the emergence of the completed transaction *just is* the emission of the β -particle, the passage of the β -particle from emitter to absorber and then the absorption of the β -particle, all together. The emergence of the transaction is *not* the emission event. Therefore the perfectly absorbing universe cannot stand in for absorber B on the left because spacetime only contains completed transactions and completed transactions are, by construction, complete four-vector particle trajectories.

Perhaps Kastner is on the right track by attempting to eliminate pseudotime from the interpretation by eliminating the dependence of the transaction mechanism on the position of *all* the possible absorbers. The resulting view of bifurcating worlds, however, is metaphysically rather strange. Indeed, in some sense there may be a correspondence between Kastner’s portrayal of bifurcating worlds and the above adoption of ‘space-pseudotime’ diagrams. There certainly seems to be a need to account for a multitude of event structures precipitated by the pseudotemporal account of the transaction mechanism. I intimated above that this counterfactual feature of the transaction mechanism is at odds with Cramer’s insistence on a real wavefunction. One might argue, if one was that way inclined metaphysically, that these facets of the interpretation can be made coherent by allowing for bifurcating worlds such as Kastner’s.

Unless one were that way inclined, however, Maudlin’s objection remains damaging. I do not, though, consequently follow Maudlin in thinking that “any theory in which both backwards and forwards influences conspire to shape events will face this same challenge”. Recall that the selection of problems introduced in §6.5 that Maudlin thought peculiar to Cramer’s theory arose as a result of the pseudotemporal account of the transaction mechanism. These problems are intimately linked, I think: the pseudotime heuristic and the reality of the wavefunction are difficult to reconcile. As we have just seen, however, a more significant worry is the underconstrained nature of the behaviour of the system. Maudlin believes this to be endemic

to retrocausal theories in general. I contend that it is the lack of causal symmetry in Cramer's theory that is to blame here.

6.8 Causal symmetry

The pseudotemporal account of the transaction mechanism that Cramer provides, while retrocausal in the sense that it contains both retarded and advanced influences, is not time symmetric. The initial offer wave always precedes (pseudotemporally) the other processes of the transaction and thus the initial conditions of a quantum system described by the transactional interpretation have primacy over any other boundary condition in constraining the dynamics. This is instrumental in rendering the pseudotemporal account of the transaction mechanism problematic. The varying event structures associated with different possible outcomes of a single stochastic event that we encountered above would not arise if the transaction mechanism endowed both the retarded and advanced elements of the transaction with equivalent causal significance. To do so would amount to constraining the transaction mechanism from both temporal ends and this, in turn, would be enough to constrain the event structure uniquely in spacetime.

Indeed, Maudlin suggests something along these lines as the key to a successful retrocausal theory:

If the course of present events depend on the future and the shape of the future is in part determined by the present then there must be some structure which guarantees the existence of a coherent mutual adjustment of all the free variables. (2002, p. 201)

Thus due to the causal asymmetry of the pseudotemporal account of the transaction mechanism, the retarded and advanced elements of Cramer's theory demonstrably do *not* have a structure which guarantees the existence of a coherent mutual adjustment of all the free variables.

Maudlin realises that this failure to provide a coherent mutual adjustment of free variables is indeed the cardinal problem of the transactional interpretation, but suggests that the reason for this is simply because it is retrocausal. According to Maudlin, in theories without retrocausation (which Maudlin, following Bell, calls 'local' theories),

solutions to the field equations at a point are constrained only by the values of quantities in one light cone (either past or future) of a point. Thus in a deterministic theory, specifying data along a hyperplane of simultaneity suffices to fix a unique solution at all times, past and future of the plane. Further, the solutions can be generated sequentially: the solution at $t = 0$ can be continued to a solution at $t = 1$ without having had to solve for any value at times beyond $t = 1$. Thus the physical state at one time generates states at all succeeding times in turn. . .

[In a stochastic theory] fixing the physical state in the back light cone of a point may not determine the physical state there, but it does determine a unique probability measure over the possible states such that events at space-like separation are statistically independent of one another. . . The present moment makes all of its random choices independently and then generates the probabilities for the immediate future, and so on. (2002, p. 201)

He continues,

Any theory with both backwards and forwards causation cannot have such a structure. Data along a single hypersurface do not suffice to fix the immediate future since that in turn may be affected by its own future. The metaphysical picture of the past generating the future must be abandoned, and along with it the mathematical tractability of local theories. (2002, p. 201)

Maudlin's argument against retrocausality can thus be construed as follows:

- (i) retrocausal theories must have a structure which guarantees the coherent mutual adjustment of free variables;
- (ii) local theories are mathematically tractable and fit a metaphysical picture of the past generating the future because solutions to the field equations require only data along a single hypersurface;
- (iii) data along a single hypersurface are not sufficient for a retrocausal theory to guarantee the coherent mutual adjustment of free variables;
- (iv) therefore, retrocausal theories must abandon both the metaphysical picture of the past generating the future and with it mathematical tractability.

There are multiple reasons to be wary of this argument, which we will address here in turn. As a starting point, let us consider the claim that retrocausal theories must have a structure which guarantees the coherent mutual adjustment of free

variables and that, because of this, retrocausal theories will be underdetermined by the data along a single hypersurface. Insofar as this is the case, the transactional interpretation can be seen as an attempt to achieve the former but with a failed mechanism for remedying the latter. The failure of the transactional interpretation to achieve this, however, is not because it is a retrocausal theory, rather, as indicated above, it is because it lacks causal symmetry in its pseudotemporal account of the fundamental quantum causal mechanism. I made the suggestion above that temporal symmetry could be achieved by the transactional interpretation if both initial and final boundary constraints were employed. It should now be clear that if such constraints were present in the formalism of a retrocausal theory then this would also debase any underdetermination claim; an increase in the available data would suffice to determine uniquely the behaviour of the system. It is not the case that such a retrocausal theory would, despite Maudlin's declaration to the contrary, elicit the abandonment of mathematical tractability; we will explore this further in just a moment. A related concern, however, is that it is not entirely clear that Maudlin's underdetermination claim should worry us in the first place.

6.8.1 Causality and determination

Consider Maudlin's reasoning that, in a retrocausal setting, data along a single hypersurface do not suffice to fix the immediate future since that in turn may be affected by its own future. If such a feature of retrocausal theories cannot guarantee the coherent mutual adjustment of free variables, then by symmetry so should the temporal reverse of this reasoning fail to guarantee the coherent mutual adjustment of free variables in ordinary forwards-in-time causal cases, i.e. data along a single hypersurface should not suffice to fix the immediate past since that in turn may be affected by its own past.⁶ This is clearly not correct. In a deterministic theory, data along a single hypersurface are sufficient to determine a unique solution to the field equations and thus determine the behaviour of the system at all times, past and future. Thus the data at some time t_0 determine not only the data at t_{-1} but also the data at t_{-2} , which is normally thought to have a causal influence on the data at t_{-1} (see Figure 6.4(i)). We see quite clearly here that the data at t_{-2} is not an independent condition of the sort that could stymie the coherent mutual adjustment

⁶See also Evans, Price and Wharton (forthcoming) for the same point.

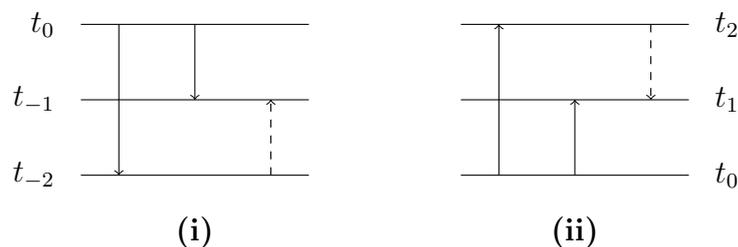


Figure 6.4: Determination and causation in (i) ordinary forwards-in-time causal theories and (ii) retrocausal theories. The black arrows indicate determination and the dashed arrows indicate what we would like to think of as causal influences in those cases.

of free variables. By the same token the data at t_0 determine not only the data at t_1 but also the data at t_2 , which in a retrocausal setting can be thought to have a causal influence on the data at t_1 (see Figure 6.4(ii)). We can now see just as clearly that the data at t_2 is likewise not an independent condition of the sort that Maudlin claims renders retrocausal theories underdetermined.

It appears as though Maudlin’s mistaken underdetermination claim emerges from a manifest tension between the temporal asymmetry of his “metaphysical picture of the past generating the future” and the temporal symmetry of determination in which “data along a hyperplane of simultaneity suffices to fix a unique solution at all times, past and future of the plane”. The tension stems from the distinctly causal notion of “generation” in Maudlin’s metaphysical picture in contrast to the “fixity” of a unique solution in his characterisation of determinism. We can alleviate this tension with the sort of carefully constructed picture of reality we developed in the last chapter (§5.4).⁷ Recall that causality can be characterised as a perspectival notion that builds upon an interventionist account of causation. Such a characterisation of causality permits us then to strike a harmony between our causal intuitions, such as deliberation, and the intuition that future events are fixed within a deterministic framework by realising that we, as spatiotemporally bound agents, are constrained in our epistemic access to the events in spacetime. With such a picture in mind, we are able to attribute both t_{-2} and t_2 with causal significance insofar as we are ignorant of the complete data at t_{-1} and t_1 respectively and also at t_0 . If we then utilise this surrogate picture of reality to reconcile causality and determinism we can

⁷See also Price and Weslake (2010) for an exposition of a similar sort of picture.

see by the above reasoning that Maudlin's argument for underdetermination loses a large part of its authority.

6.8.2 A tractable alternative

The details of this surrogate picture also arise in the context of another of Maudlin's claims. Let us for argument's sake grant that the underdetermination problem of retrocausal theories must be remedied and return to Maudlin's argument that doing so renders these theories mathematically intractable. Consider the Schrödinger equation, the wave equation of nonrelativistic quantum mechanics: it is an example of an equation that requires only data along a single hypersurface to fix a unique solution. This is because it is first-order in time. If we consider the Klein-Gordon equation, which is second-order in time, we see that it requires twice the initial data to determine a solution. Ordinarily solutions to this classical scalar field equation are found by imposing two independent initial boundary conditions: the solution to the field equations at some particular time as well as the first time derivative of this solution. According to Wharton (2010), the Klein-Gordon equation has resisted interpretation as a relativistic quantum mechanical wave equation partly due to this increase in required initial data. Wharton makes the suggestion that a time symmetric approach to quantum mechanics can provide a natural resolution to this interpretational obstacle. Rather than imposing two independent initial boundary conditions on solutions to the Klein-Gordon equation, one can impose two boundary conditions *at two different times*. This can be interpreted as supplying the field equations with data along two different instantaneous hypersurfaces or, likewise, as determining the behaviour of any system described by the Klein-Gordon equation with initial and final constraints. This causal symmetry, of course, is just the suggestion made above for overcoming Maudlin's underdetermination claim targeting retrocausal theories.⁸

Within Wharton's time symmetric scheme the full solution to the field equations cannot be known before any final constraint becomes epistemically accessible. The initial and final boundary conditions can be pictured as representative of consecutive external measurements on some quantum system. Without knowledge of the later

⁸As well as Wharton (2010), see also Sutherland (2008) for an example of a retrocausal theory with a symmetric causal structure.

measurement to be performed on the system one cannot solve the field equations and thus one cannot know the exact state of the system between the measurements. What one can know before the later measurement, however, is some best approximation to the full solution based on the initial data and it seems reasonable to think this would be the ordinary wavefunction of the Schrödinger equation. This then yields a ‘hidden variable’ theory of sorts where the wavefunction of the quantum formalism is interpreted as representing an observer’s knowledge of the system and the full solution to the Klein-Gordon equation is interpreted as representing the actual state of the system, hidden from the observer. Upon measuring the system the observer gains knowledge of the final constraint and can retrodict the intervening state based on the now attainable full solution to the Klein-Gordon equation. This careful attention to the epistemic limitations of the spatiotemporally bound observer is just the same principle that buttresses our surrogate picture from Chapter 5.

Moreover, such a retrocausal scheme for modelling quantum processes seems by no means ‘mathematically intractable’; on the contrary, not only do we have a straightforward algorithm for calculating the properties of any particular quantum system but we also have a clear metaphysical prescription, whose limitations reflect our limitations as spatiotemporally bound observers, for representing this system. We have seen that the metaphysical picture that Maudlin ties to mathematical tractability, that of the past generating the future, is abandoned trivially within any retrocausal scheme but this evidently does not imply that we must also abandon mathematical tractability. Indeed, by emphasising this traditionalist metaphysical picture of reality, it seems as though what Maudlin has in mind when he says ‘mathematical tractability’ is something commensurate with a particular form of initial value problem. Retrocausal theories of quantum mechanics aside, if we look toward some of our more established physical theories we see that there is little justification for this characterisation of mathematical tractability.

6.8.3 Classical tractability

In the first place, the representation of the dynamical behaviour of classical physical systems according to analytical mechanics certainly does not preclude all but an initial value metaphysics. Granted, the Hamiltonian formulation of dynamics appears to provide good support for this metaphysical picture: the dynamical arena

of Hamiltonian mechanics, phase space, is a space of possible initial values with a geometric structure that allows the determination of a unique dynamical path given any single point in the space (recall Chapter 1). However, when one considers how this geometric structure is derived from the formalism of analytical mechanics, one finds that this Hamiltonian picture is merely (as Lanczos (1970) points out) a “remarkable simplification” of a deeper dynamical picture. The geometric structure of phase space is encoded in Hamilton’s equations of motion and, according to Lanczos, there are two ways that these equations can be derived. The first way is to decompose the second-order Euler-Lagrange equations of motion (1.4) into two first-order equations that can be transformed into Hamilton’s equations by application of a Legendre transformation (1.9). The Euler-Lagrange equations themselves are attained by way of the variational principle: the equations of motion are the necessary and sufficient conditions for the action integral to remain stationary under arbitrary variations of the configuration of the system *given the initial and final configurations of the system*. Thus it would seem that the Hamiltonian formulation might indeed be built upon temporally symmetric boundary conditions.

The second way of deriving Hamilton’s equations of motion, however, observes that since the Legendre transformations are completely symmetric there is no requirement that we must take the Euler-Lagrange formulation of mechanics as primary. As such, one can formulate Hamilton’s equations directly without the Lagrangian equations nor the Legendre transformations (Lanczos, 1970, p. 169); recall §1.4. However, to do so one must produce a new action integral in terms of an extended set of independent variables and subject it once again to a variational principle (1.10); Hamilton’s equations become the conditions for a stationary action integral under arbitrary variations which are again constrained by *initial and final boundary conditions*. Regardless then of how one constructs the geometry of Hamiltonian phase space, the fundamental element of analytical mechanics remains the specification of initial and final boundary conditions as part of the variational principle. Thus it appears as unlikely that one might find justification for Maudlin’s characterisation of ‘mathematical tractability’ in analytical mechanics.

The case of general relativity is not so clear cut. On the one hand, it is more than reasonable to take the central lesson of general relativity to be that the fundamental ontological unit of our reality is a four dimensional solution to Einstein’s field equations; solutions are clearly not obtained in Maudlin’s ‘mathematically tractable’

way.⁹ On the other hand, though, considerable effort has been spent over the last half a century attempting to cast general relativity in a form that explicitly separates out a single temporal dimension from three spatial dimensions¹⁰, which would appear the best hope for a justification of Maudlin’s metaphysical picture: as long as a spacetime is globally hyperbolic, a solution can be generated from data on any Cauchy surface. However, a new difficulty arises herein: due to the foliation invariance of such formulations of general relativity there exists a troublesome indeterminacy problem. Not only does specification of a single 3-geometry (of which a phase space point is comprised when combined with the relevant canonically conjugate momentum variable) fail to determine uniquely a dynamical path but, as Pooley (2001) points out, “the specification of an *initial sequence* of 3-geometries is not sufficient to allow us to predict which continuation of the sequence will be actualized” (emphasis added). This indeterminacy may not be as pernicious as it first appears, since it is a function of gauge freedom, and thus every actualised sequence represents the same spacetime sliced in different ways. However, at the level of hypersurfaces, data along a single hypersurface is insufficient to determine a unique continuation in its immediate future.¹¹ It is up for grabs then whether or not general relativity fits Maudlin’s characterisation of ‘mathematical tractability’.

6.9 Cramer’s missing structure and Maudlin’s misdirected metaphysics

Maudlin’s inventive thought experiment exposes a deep problem within Cramer’s theory: the causal structure of the transaction mechanism cannot constrain uniquely and consistently the behaviour of particular quantum systems. I claim that what the transactional interpretation is missing is a causally symmetric account of the transaction mechanism: that is, both initial and final boundary constraints with equal causal significance influencing the dynamics of the system. Such a causally symmetric mechanism would serve to ensure the coherent mutual adjustment of all the relevant free variables. In contrast, Maudlin attributes this shortcoming of the transactional interpretation to the inability of a retrocausal theory to supply a structure that could achieve such mutual adjustment of variables. Moreover,

⁹See Brown (2005, §9.2.2) for a discussion of this point.

¹⁰See, for instance, Dirac (1958), Bergmann (1961), Arnowitt *et al.* (1962) and Barbour (1994a).

¹¹This is related to the thin and thick sandwich problems; see Baierlein, Sharp and Wheeler (1962) and Wheeler (1964).

Maudlin claims that the inability of retrocausal theories to achieve this is due to a fundamental incongruence between retrocausality and his “metaphysical picture of the past generating the future”. This picture underpins his underdetermination challenge to retrocausal theories and his notion of mathematical tractability.

I hope to have shown that we have good reason to be wary of Maudlin’s metaphysical picture and its connection to mathematical tractability. Firstly, Maudlin’s underdetermination challenge to retrocausality can be subverted if one is careful to spell out the metaphysical difference between causality and determination; Maudlin’s picture evidently does not achieve this. Secondly, we saw an example of a retrocausal theory of quantum mechanics that does not encounter any problems with mathematical tractability, despite not adhering to the edict of Maudlin’s picture. Thirdly, it seems unlikely that analytical mechanics, and possibly general relativity, can be used to support an initial value metaphysics, despite being arguably the best place to begin looking for mathematical tractability in physical theories. At the very least we can conclude from this that Maudlin’s picture cannot be used as strongly as he may have liked as an objection against retrocausality.

The transactional interpretation, and Maudlin’s critique, does show us something important: for retrocausality to be taken seriously in contemporary physics, it must be supported by a coherent picture of reality and, above all, this picture would do well to be causally symmetric.

Summary

Overview

I set out at the beginning of this thesis to explore how time is portrayed within our modern physical theories. This exploration has not produced a linear narrative; on the contrary, the manner in which we have traversed a network of overlapping ideas in mathematics, physics and philosophy demonstrates just how centrally a holistic study of time fits within the discipline of philosophy of physics. I adopted from the outset a clear statement of methodology in which modern science is treated as the authority on, and primary guide to, the nature of time and any metaphysical inquiry is considered legitimate only when motivated exclusively by contemporary science. Rather than acting to the detriment of philosophical investigation, these clear guidelines have enabled a precise and rich discussion of the issues that we face in the philosophy of time.

Part I was a demonstration of this methodology: of what an analysis of time amounts to when we take seriously the doctrine that modern physics should be treated as the primary guide to the nature of time. Chapter 1 showed how the Newtonian picture of time arises from Newtonian mechanics and, despite the novel and interesting pictures of reality that emerge in the context of both Lagrangian and Hamiltonian mechanics, the Newtonian picture of time remains a fixture of classical mechanics. In Chapter 2 I outlined the constraints that relativity theory imposes on the traditional metaphysical debate concerning the nature of time, if this debate is to remain motivated exclusively by contemporary science. I then considered in Chapter 3 a claim of Barbour's that a timeless picture of reality arises in the context of his Machian formulation of general relativity and his interpretation of canonical quantum gravity. An interesting theme of the analysis of Part I is the range of different pictures of time that we find arising in the context of each new physical theory.

In Part II I explored a confusion that can be seen as arising due to the absence

of the methodology of Part I within the interpretation of nonrelativistic quantum mechanics: study into the nature of time should be *guided* by modern physics and thus we should be careful not to *insert* a preconceived Newtonian conception of time unwittingly into our interpretation of the quantum mechanical formalism. To this end, in Chapter 4 I introduced the hypothesis of retrocausality in quantum mechanics as a solution to the interpretational difficulties derived from Bell's theorem with a view to demonstrating that an overly Newtonian conception of time might be contributing to these difficulties. Chapter 5 stands as an independent defence of retrocausality by way of the development of a coherent picture of reality that cannot be precluded by contemporary physics on analytic grounds; in other words, the picture respects the authority of contemporary science. I employ this picture to argue in Chapter 6 that Maudlin's objection to Cramer's transactional interpretation of quantum mechanics is misguided by his insistence on an overly Newtonian conception of time; my essential claim is that Maudlin has overstated his own picture of reality.

A take-home message

I have emphasised that latent assumptions based on our intuitions about the nature of time are creating difficulties for scientific progress. Compounding the problematic nature of this issue is the close correspondence between the Newtonian picture of time and many of our intuitions about the nature of time; one could see this correspondence as responsible for providing a misleading justification for maintaining these intuitions. Indeed, the difficulties associated with nonlocality and action-at-a-distance in quantum mechanics find their root in the generative picture of Newtonian time. As far as this is problematic, nonrelativistic quantum mechanics is calling out for a fresh conceptualisation of time that is consistent with the picture of reality that arises in the context of the quantum formalism.

The picture of reality that arises from the philosophical considerations of retrocausality in quantum mechanics provides one such solution, and we have seen in Chapter 5 an independent argument for why this solution is not ruled out as a possibility on analytic grounds. What is really quite significant here, and the major result for which this work provides evidence, is that this notion of time can already be found in a particular formulation of classical mechanics: namely, the conceptual

schema of Lagrangian mechanics. The essential feature of this Lagrangian conceptual schema is that it is more naturally interpreted as supporting a teleological picture of determination (as opposed to the generative picture of Newtonian time): the determination of dynamical behaviour requires both initial and final boundary conditions. What I have demonstrated (especially through the considerations of §6.8) is that it is exactly this teleological determination that underpins an understanding of retrocausality as a physical phenomenon.

Bibliography

- Argaman, N. (2008). Bell's Theorem and the Causal Arrow of Time. [arXiv:0807.2041v2](https://arxiv.org/abs/0807.2041v2) [quant-ph].
- Arnowitt, R., Deser, S. and Misner, C. W. (1962). The Dynamics of General Relativity. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*, John Wiley & Sons Inc., New York, chapter 7, pp. 227–265.
- Arthur, R. T. W. (1995). Newton's Fluxions and Equably Flowing Time. *Stud. Hist. Phil. Sci.* **26**: 323–351. doi:10.1016/0039-3681(94)00037-A.
- Aspect, A., Delibard, J. and Roger, G. (1982a). Experimental Tests of Bell's Inequalities Using Time-Varying Analyzers. *Phys. Rev. Lett.* **49**(25): 1804–1807. doi:10.1103/PhysRevLett.49.1804.
- Aspect, A., Grangier, P. and Roger, G. (1982b). Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Phys. Rev. Lett.* **49**(2): 91–94. doi:10.1103/PhysRevLett.49.91.
- Bacciagaluppi, G. and Valentini, A. (2009). *Quantum Theory at the Crossroads*. Cambridge University Press, Cambridge.
- Baierlein, R. F., Sharp, D. H. and Wheeler, J. A. (1962). Three-Dimensional Geometry as a Carrier of Information about Time. *Phys. Rev.* **126**: 1864–1865. doi:10.1103/PhysRev.126.1864.
- Barbour, J. (1994a). The timelessness of quantum gravity: I. The evidence from the classical theory. *Class. Quant. Grav.* (11): 2853–2873. doi:10.1088/0264-9381/11/12/005.
- (1994b). The timelessness of quantum gravity: II. The appearance of dynamics in static configurations. *Class. Quant. Grav.* (11): 2875–2897. doi:10.1088/0264-9381/11/12/006.

- (1999). *The End of Time: The Next Revolution in Our Understanding of the Universe*. Oxford University Press, Oxford.
- Baron, S., Evans, P. W. and Miller, K. (2010). From Timeless Physical Theory to Timelessness. *Humana.Mente Journal of Philosophical Studies* **13**: 35–60.
- Bell, J. S. (2004). On the Einstein-Podolsky-Rosen paradox. In M. Bell (ed.), *Speakable and Unsayable in Quantum Mechanics*, Cambridge University Press, Cambridge, chapter 2, pp. 14–21.
- Belot, G. (2007). The Representation of Time and Change in Mechanics. In J. Butterfield and J. Earman (eds.), *Philosophy of Physics*, North-Holland, Amsterdam, pp. 133–227.
- Bergmann, P. G. (1949). Non-Linear Field Theories. *Phys. Rev.* **75**: 680–685. doi:10.1103/PhysRev.75.680.
- (1961). Observables in General Relativity. *Rev. Mod. Phys.* **33**: 510–514. doi:10.1103/RevModPhys.33.510.
- Berkovitz, J. (2002). On Causal Loops in the Quantum Realm. In T. Placek and J. Butterfield (eds.), *Non-locality and Modality*, Kluwer, Dordrecht, pp. 235–257.
- (2008). On predictions in retro-causal interpretations of quantum mechanics. *Stud. Hist. Phil. Mod. Phys.* **39**: 709–735. doi:10.1016/j.shpsb.2008.08.002.
- Bohm, D. (1952a). A Suggested Interpretation of the Quantum Theory in Terms of “Hidden” Variables. I. *Phys. Rev.* **85**(2): 166–179. doi:10.1103/PhysRev.85.166.
- (1952b). *Quantum Theory*. Prentice Hall, New York.
- Bohm, D. and Aharonov, Y. (1957). Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.* **108**(4): 1070–1076. doi:10.1103/PhysRev.108.1070.
- Boyer, C. B. (1970). The History of the Calculus. *The Two-Year College Mathematics Journal* **1**: 60–86.
- Brown, H. R. (2005). *Physical Relativity*. Clarendon Press, Oxford.

-
- Butterfield, J. (1992). Bell's Theorem: What it Takes. *Brit. J. Phil. Sci.* **43**: 41–83.
- (2002). Critical notice of Julian Barbour “The End of Time: The Next Revolution in Our Understanding of the Universe”. *Brit. J. Phil. Sci.* **53**: 289–330. doi:10.1093/bjps/53.2.289.
- Butterfield, J. and Isham, C. J. (1999). On the Emergence of Time in Quantum Gravity. In J. Butterfield (ed.), *The Arguments of Time*, Oxford University Press, Oxford, chapter 6, pp. 111–168. arXiv:gr-qc/9901024v1.
- Callender, C. (2006). Time in Physics. In D. M. Borchert (ed.), *Encyclopedia of Philosophy*, MacMillan Reference USA, Detroit, volume 9, pp. 493–501.
- Christian, J. (2007). Disproof of Bell's Theorem: Further Consolidations. arXiv:0707.1333v2 [quant-ph].
- Costa de Beauregard, O. (1953). Mécanique Quantique. *Comptes Rendus de l'Académie des Sciences* **T236**: 1632–1634.
- (1976). Time Symmetry and Interpretation of Quantum Mechanics. *Found. Phys.* **6**: 539–559. doi:10.1007/BF00715107.
- (1977). Time symmetry and the Einstein paradox. *Il Nuovo Cimento* **42**: 41–63. doi:10.1007/BF02906749.
- Cramer, J. G. (1980). Generalized absorber theory and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. D* **22**: 362–676. doi:10.1103/PhysRevD.22.362.
- (1986). The transactional interpretation of quantum mechanics. *Rev. Mod. Phys.* **58**: 647–687. doi:10.1103/RevModPhys.58.647.
- (1988). An Overview of the Transactional Interpretation of Quantum Mechanics. *Int. J. Theor. Phys.* **27**: 227–236. doi:10.1007/BF00670751.
- Dainton, B. (2001). *Time and Space*. Acumen Publishing Limited, Chesham.
- DeWitt, B. S. (1962). The quantization of geometry. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*, John Wiley & Sons Inc., New York, chapter 8, pp. 266–381.

- Dickson, W. M. (1998). *Quantum chance and non-locality*. Cambridge University Press, Cambridge.
- Dieks, D. (1991). Time in special relativity and its philosophical significance. *Eur. J. Phil.* **12**: 253–259. doi:10.1088/0143-0807/12/6/002.
- (2006). Becoming, relativity and locality. In D. Dieks (ed.), *The Ontology of Spacetime*, Elsevier, Amsterdam, pp. 157–176. doi:10.1016/S1871-1774(06)01008-4.
- Dirac, P. A. M. (1930). *The Principles of Quantum Mechanics*. Oxford University Press, London.
- (1938). Classical Theory of Radiating Electrons. *Proc. R. Soc. A* **167**: 148–169.
- (1958). The Theory of Gravitation in Hamiltonian Form. *Proc. R. Soc. A* **246**: 333–343. doi:10.1098/rspa.1958.0142.
- (1964). *Lectures on Quantum Mechanics*. Yeshiva University, New York.
- Dummett, M. (1964). Bringing About the Past. *The Philosophical Review* **73**(3): 338–359.
- Earman, J. (1986). *A Primer on Determinism*. D. Reidel Publishing Company, Dordrecht.
- Einstein, A. (1952). On the Electrodynamics of Moving Bodies. In *The Principle of Relativity*. Trans. Perrett, W. and Jeffery, G. B., Dover Publications, New York, chapter 3, pp. 35–65.
- Einstein, A., Podolsky, B. and Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* **47**: 777–780. doi:10.1103/PhysRev.47.777.
- Ellis, G. F. R. (2007). Physics in the Real Universe: Time and Spacetime. In V. Petkov (ed.), *Relativity and the Dimensionality of the World*, Springer, Netherlands, volume 153, pp. 49–79. doi:10.1007/978-1-4020-6318-3_4.

-
- Evans, P. W., Price, H. and Wharton, K. B. (forthcoming). New Slant on the EPR-Bell Experiment. *Brit. J. Phil. Sci.* arXiv:1001.5057v3 [quant-ph].
- French, S. (1989). Identity and Individuality in Classical and Quantum Physics. *Australasian Journal of Philosophy* **67**: 432–446. doi:10.1080/00048408912343951.
- (1998). On The Withering Away of Physical Objects. In E. Castellani (ed.), *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, Princeton University Press, Princeton, chapter 6, pp. 93–113.
- French, S. and Ladyman, J. (2003). Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure. *Synthese* **136**: 31–56. doi:10.1023/A:1024156116636.
- Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton University Press, New Jersey.
- Geroch, R. (1970). Domain of Dependence. *J. Math. Phys.* **11**: 437–449. doi:10.1063/1.1665157.
- Gödel, K. (1949). An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation. *Rev. Mod. Phys.* **21**: 447–450. doi:10.1103/RevModPhys.21.447.
- Gryb, S. (2010). Jacobi's Principle and the Disappearance of Time. arXiv:0804.2900v3 [gr-qc].
- Hamilton, W. R. (1834). On the General Method in Dynamics. *Phil. Trans. R. Soc.* **124**: 247–308.
- Healey, R. (2002). Can Physics Coherently Deny the Reality of Time? In C. Callender (ed.), *Time, Reality and Experience*, Cambridge University Press, Cambridge, pp. 293–316.
- Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik* **33**: 879–893. doi:10.1007/BF01328377.

- (1955). The Development of the Interpretation of the Quantum Theory. *In* W. Pauli (ed.), *Niels Bohr and the Development of Physics*, Pergamon, London, pp. 12–29.
- Hinchliff, M. (1996). The Puzzle of Change. *Noûs* **30**: 119–136.
- Hokkyo, N. (1988). Variational formulation of transactional and related interpretations of quantum mechanics. *Found. Phys. Lett.* **1**: 293–299. doi:10.1007/BF00690070.
- Howard, D. (2004). Who Invented the “Copenhagen Interpretation”? A Study in Mythology. *Phil. Sci.* **71**: 669–682.
- Hughes, R. I. G. (1989). *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press, Cambridge, Massachusetts.
- Ismael, J. (2002). Rememberances, Momentos, and Time-Capsules. *In* C. Callender (ed.), *Time, Reality and Experience*, Cambridge University Press, Cambridge, pp. 317–328.
- Janis, A. L. (1983). Simultaneity and Conventionality. *In* R. S. Cohen and L. Laudan (eds.), *Physics, Philosophy and Psychoanalysis*, D. Reidel Publishing Company, Dordrecht, pp. 101–110.
- Kastner, R. (2006). Cramer’s Transactional Interpretation and Causal Loop Problems. *Synthese* **150**: 1–14. doi:10.1007/s11229-004-6264-9.
- (2010). The Quantum Liar Experiment in Cramer’s Transactional Interpretation. *Stud. Hist. Phil. Mod. Phys.* **41**: 86–92. arXiv:0906.1626v5 [quant-ph].
- Kastner, R. and Cramer, J. G. (2010). Why Everettians Should Appreciate the Transactional Interpretation. arXiv:1001.2867v3 [quant-ph].
- Kroes, P. (1985). *Time: its structure and role in physical theories*. D. Reidel Publishing Company, Dordrecht.
- Ladyman, J. (1998). What is Structural Realism? *Stud. Hist. Phil. Sci.* **29**: 409–424. doi:10.1016/S0039-3681(98)80129-5.

-
- Ladyman, J. and Ross, D. (2007). *Every Thing Must Go*. Oxford University Press, Oxford.
- Lagrange, J. L. (1853). *Mécanique Analytique*. Mallet-Bachelier, Paris.
- Lanczos, C. (1970). *The Variational Principles of Mechanics*. Dover Publications, New York.
- Le Poidevin, R. (1998). The Past, Present, and Future of the Debate about Tense. In R. Le Poidevin (ed.), *Questions of Time and Tense*, Clarendon Press, Oxford, pp. 13–42.
- Lewis, G. N. (1926). The Nature of Light. *Proc. Nat. Acad. Sci.* **12**: 22–29.
- Mach, E. (1960). *The Science of Mechanics: A Critical and Historical Account of its Development*. Trans. McCormack, T. J. Open Court Publishing Company, LaSalle, Illinois.
- Malament, D. B. (1977). Causal Theories of Time and the Conventionality of Simultaneity. *Noûs* **11**: 293–300.
- (2007). Classical Relativity Theory. In J. Butterfield and J. Earman (eds.), *Philosophy of Physics*, North-Holland, Amsterdam, pp. 229–273.
- Marchildon, L. (2006). Causal Loops and Collapse in the Transactional Interpretation of Quantum Mechanics. *Physics Essays* **10**: 422–429. arXiv:quant-ph/0603018v2.
- Maudlin, T. (2002). *Quantum Non-Localilty and Relativity*. Blackwell Publishing, Oxford.
- Maxwell, N. (1985). Are Probabilism and Special Relativity Incompatible? *Phil. Sci.* **52**: 23–43.
- McCall, S. (1976). Objective Time Flow. *Phil. Sci.* (43): 337–362.
- (2001). Time Flow. In L. N. Oaklander (ed.), *The Importance of Time*, Kluwer Academic Publishers, Dordrecht, chapter 11, pp. 143–151.

- McTaggart, J. E. (1908). The Unreality of Time. *Mind* **17**: 457–474. doi:10.1093/mind/XVII.4.457.
- Miller, D. J. (1996). Realism and time symmetry in quantum mechanics. *Phys. Lett. A* **222**: 31–36. doi:10.1016/0375-9601(96)00620-2.
- (1997). Conditional probabilities in quantum mechanics from time-symmetric formulation. *Il Nuovo Cimento* **112B**: 1577–1592.
- Minkowski, H. (1952). Space and Time. In *The Principle of Relativity*. Trans. Perrett, W. and Jeffery, G. B., Dover Publications, New York, chapter 5, pp. 73–91.
- Mott, N. F. (1929). The Wave Mechanics of α -Ray Tracks. *Proc. R. Soc. A* **126**: 79–84.
- Newton, I. (1962). *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his System of the World*. Trans. Motte, A., revised Cajori, F. University of California Press, Berkeley.
- Peres, A. (1962). On Cauchy's Problem in General Relativity. *Il Nuovo Cimento* **26**: 53–62. doi:10.1007/BF02754342.
- Pirani, F. A. E. and Schild, A. (1950). On the Quantization of Einstein's Gravitational Field Equations. *Phys. Rev.* **79**: 986–991. doi:10.1103/PhysRev.79.986.
- Pooley, O. (2001). Relationism Rehabilitated? II: Relativity. <http://philsci-archive.pitt.edu/221/>.
- Price, H. (1984). The philosophy and physics of affecting the past. *Synthese* **61**: 299–324. doi:10.1007/BF00485056.
- (1991). The Asymmetry of Radiation: Reinterpreting the Wheeler-Feynman Argument. *Found. Phys.* **21**(8): 959–975. doi:10.1007/BF00733218.
- (1994). A Neglected Route to Realism about Quantum Mechanics. *Mind* **103**: 303–336. doi:10.1093/mind/103.411.303.
- (1996). *Time's Arrow and Archimedes' Point*. Oxford University Press, New York.

-
- (1997). Time Symmetry in Microphysics. *Phil. Sci.* **64**: 235–244. arXiv:quant-ph/9610036v1.
- (2001). Backwards causation, hidden variables, and the meaning of completeness. *Pramana J. Phys.* **56**: 199–209. doi:10.1007/s12043-001-0117-6.
- (2007). Causal Perspectivalism. In H. Price and R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford University Press, Oxford, chapter 10, pp. 250–292.
- (2008). Toy models for retrocausality. *Stud. Hist. Phil. Mod. Phys.* **39**: 752–776. doi:10.1016/j.shpsb.2008.05.006.
- (2010). Does Time-Symmetry Imply Retrocausality: How the Quantum World Says "Maybe". arXiv:1002.0906v1 [quant-ph].
- (forthcoming). The Flow of Time. In C. Callender (ed.), *The Oxford Handbook of Time*, Oxford University Press, Oxford.
- Price, H. and Weslake, B. (2010). The Time-Asymmetry of Causation. In H. Beebe, C. Hitchcock and P. Menzies (eds.), *The Oxford Handbook of Causation*, Oxford University Press, New York, chapter 20, pp. 414–443.
- Pullin, J. (2003). Canonical quantization of general relativity: the last 18 years in a nutshell. *AIP Conference Proceedings* **668**: 141–153. doi:10.1063/1.1587095. arXiv:gr-qc/0209008v1.
- Putnam, H. (1967). Time and Physical Geometry. *Journal of Philosophy* **64**: 240–247.
- Quine, W. V. O. (1951). Ontology and Ideology. *Phil. Stud.* **2**: 11–15. doi:10.1007/BF02198233.
- Rickles, D. (2006). Time and Structure in Canonical Gravity. In S. French, D. Rickles and J. Staasi (eds.), *Structural Foundations of Quantum Gravity*, Oxford University Press, Oxford, chapter 6, pp. 152–195.
- (2008). Quantum Gravity: A Primer for Philosophers. In D. Rickles (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, Ashgate Publishing Limited, Aldershot, chapter 5, pp. 262–365.

- Rietdijk, C. W. (1966). A rigorous proof of determinism derived from the special theory of relativity. *Phil. Sci.* **33**: 341–344.
- (1978). Proof of a retroactive influence. *Found. Phys.* **8**: 615–628. doi:10.1007/BF00717585.
- Rodrigues, W. A. J., de Souza, Q. A. G. and Bozhkov, Y. (1995). The Mathematical Structure of Newtonian Spacetime: Classical Dynamics and Gravitation. *Found. Phys.* **25**: 871–924. doi:10.1007/BF02080568.
- Rovelli, C. (1995). Analysis of the Distinct Meanings of the Notion of Time in Different Physical Theories. *Il Nuovo Cimento* **110**: 81–93. doi:10.1007/BF02741291.
- (2004). *Quantum Gravity*. Cambridge University Press, Cambridge.
- Saunders, S. (2002). How Relativity Contradicts Presentism. In C. Callender (ed.), *Time, Reality and Experience*, Cambridge University Press, Cambridge, pp. 277–292.
- Savitt, S. (forthcoming). Time in the Special Theory of Relativity. In C. Callender (ed.), *The Oxford Handbook of Time*, Oxford University Press, Oxford.
- Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Ann. Phys.* **79**: 361–376. doi:10.1002/andp.19263851302.
- Schutz, B. F. (1980). *Geometrical Methods of Mathematical Physics*. Cambridge University Press, Cambridge.
- Sellars, W. (1963). *Science, Perception and Reality*. Routledge & Kegan Paul, London.
- Sklar, L. (1977). *Space, Time, and Spacetime*. University of California Press, Berkeley.
- Stein, H. (1968). On Einstein-Minkowski Space-time. *Journal of Philosophy* **65**: 5–23.
- Sutherland, R. I. (1983). Bell’s theorem and backwards-in-time causality. *Int. J. Theor. Phys.* **22**: 377–384. doi:10.1007/BF02082904.

-
- (1998). Density Formalism for Quantum Theory. *Found. Phys.* **28**: 1157–1190. doi:10.1023/A:1018850120826.
- (2008). Causally symmetric Bohm model. *Stud. Hist. Phil. Mod. Phys.* **39**: 782–805. doi:10.1016/j.shpsb.2008.04.004.
- Tetrode, H. (1922). Über den Wirkungszusammenhang der Welt. Eine Erweiterung der klassischen Dynamik. *Zeitschrift für Physik* **10**: 317–328. doi:10.1007/BF01332574.
- Tooley, M. (2000). *Time, Tense and Causation*. Oxford University Press, Oxford.
- van Fraassen, B. (1991). *Quantum Mechanics: An Empiricist View*. Oxford University Press, Oxford.
- von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin.
- Westman, H. and Sonogo, S. (2009). Coordinates, observables and symmetry in relativity. *Annals of Physics* **324**: 1585–1661. doi:10.1016/j.aop.2009.03.014.
- Wharton, K. B. (2007). Time-Symmetric Quantum Mechanics. *Found. Phys.* **37**: 159–168. doi:10.1007/s10701-006-9089-1.
- (2010). A Novel Interpretation of the Klein-Gordon Equation. *Found. Phys.* **40**: 313–332. doi:10.1007/s10701-009-9398-2.
- Wheeler, J. A. (1964). Geometrodynamics and the Issue of the Final State. In B. DeWitt and C. DeWitt (eds.), *Relativity, Groups and Topology: 1963 Les Houches Lectures*, Gordon and Breach, New York, pp. 315–320.
- Wheeler, J. A. and Feynman, R. P. (1945). Interaction with the Absorber as the Mechanism of Radiation. *Rev. Mod. Phys.* **17**: 157–181. doi:10.1103/RevModPhys.17.157.
- Wigner, E. P. (1970). On Hidden Variables and Quantum Mechanical Probabilities. *Am. J. Phys.* **38**: 1005–1009. doi:10.1119/1.1976526.
- Williams, D. C. (1951). The Myth of Passage. *Journal of Philosophy* **48**: 457–472.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, New York.

Wüthrich, C. (2010). No presentism in quantum gravity. In V. Petkov (ed.), *Space, Time, and Spacetime*, Springer, New York. doi:10.1007/978-3-642-13538-5_12.

Zimmerman, D. (2008). The Privileged Present: Defending an ‘A-Theory’ of Time. In T. Sider, J. Hawthorne and D. Zimmerman (eds.), *Contemporary Debates in Metaphysics*, Blackwell, Oxford, pp. 211–225.

——— (forthcoming). Presentism and the Space-Time Manifold. In C. Callender (ed.), *The Oxford Handbook of Time*, Oxford University Press, Oxford. [http://fas-philosophy.rutgers.edu/zimmerman/Presentism and Rel.for.Web.2.pdf](http://fas-philosophy.rutgers.edu/zimmerman/Presentism%20and%20Rel.for.Web.2.pdf).