SPACETIME SYMMETRIES AND THE CPT THEOREM

BY HILARY GREAVES

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Frank Arntzenius

and approved by

New Brunswick, New Jersey

May, 2008

ABSTRACT OF THE DISSERTATION

Spacetime symmetries and the CPT theorem

by Hilary Greaves

Dissertation Director: Frank Arntzenius

This dissertation explores several issues related to the CPT theorem.

Chapter 2 explores the meaning of spacetime symmetries in general and time reversal in particular. It is proposed that a third conception of time reversal, 'geometric time reversal', is more appropriate for certain theoretical purposes than the existing 'active' and 'passive' conceptions. It is argued that, in the case of classical electromagnetism, a particular nonstandard time reversal operation is at least as defensible as the standard view. This unorthodox time reversal operation is of interest because it is the classical counterpart of a view according to which the so-called 'CPT theorem' of quantum field theory is better called 'PT theorem'; on this view, a puzzle about how an operation as apparently non-spatio-temporal as charge conjugation can be linked to spacetime symmetries in as intimate a way as a CPT theorem would seem to suggest dissolves.

In chapter 3, we turn to the question of whether the CPT theorem is an essentially quantum-theoretic result. We state and prove a classical analogue of the CPT theorem for systems of tensor fields. This classical analogue, however,

appears not to extend to systems of spinor fields. The intriguing answer to our question thus appears to be that the CPT theorem *for spinors* is essentially quantum-theoretic, but that the CPT theorem for tensor fields applies equally to the classical and quantum cases.

Chapter 4 explores a puzzle that arises when one puts the CPT theorem alongside a standard way of understanding spacetime symmetries, according to which (latter) spacetime symmetries are to be understood in terms of background spacetime structure. The puzzle is that a 'PT theorem' amounts to a statement that the theory may not make essential use of a preferred direction of time, and this seems odd. We propose a solution to that puzzle for the case of tensor field theories.

Acknowledgements

Each of the main chapters in this dissertation is, or is a very slightly adapted version of, a research article prepared for submission to a professional journal. The articles in chapters 2 and 4 are, at the time of writing, being considered for publication by *The British Journal for the Philosophy of Science*. That in chapter 3 has not yet been submitted.

Chapter 2 was written jointly by Frank Arntzenius and myself. A draft containing ancestors of much of sections 2.2, 2.3 and 2.4 was produced by Arntzenius, at that time working alone, in July 2005. These have been significantly rewritten and added to (by both of us) following discussion between the two of us. Section 2.5 (on structuralism) was written by me. We estimate that our overall contributions to this paper are equal.

Chapters 3 and 4 are single-authored.

I would like to thank Robert Geroch for extensive correspondence on the 'classical PT theorem' proved in chapter 3, and an anonymous referee for *The British Journal for the Philosophy of Science* for extremely helpful feedback on the paper reproduced here as chapter 4.

Table of Contents

Chapter 1

Introduction

This dissertation consists of three papers in the philosophy of physics.

All three began as attempts to understand the CPT theorem of quantum field theory. This theorem states that any relativistic (that is, Lorentz-invariant) quantum field theory must also be invariant under CPT, the composition of charge conjugation, parity reversal and time reversal. Two things seem initially puzzling about such a theorem:

- How can there be such an intimate relationship between *spatiotemporal* symmetries (Lorentz invariance, parity reversal, time reversal) on the one hand, and *charge conjugation*, not obviously a spatiotemporal notion at all, on the other?

- How can it come about that one symmetry (e.g. Lorentz invariance) entails another (e.g. CPT) *at all*?

These two types of puzzlement are quite distinct. The first paper, 'Time reversal in classical electromagnetism' (chapter 2), began as an attempt to solve the first puzzle. The underlying thought here is that, while the *mathematics* that goes into proving that such-and-such a transformation on quantum fields is (for present purposes) beyond doubt, one calls the theorem in question a 'CPT theorem' only on the basis of certain assumptions about which transformations on sets of fields are to be interpreted as implementing time reversal, charge conjugation and parity reversal, and that these latter assumptions are *not* beyond question. In fact, they have occasionally been questioned. John S Bell, for example, notes that while the transformation for which we prove a theorem is naturally interpreted as 'CPT' under a particle interpretation of quantum field theory, it is more naturally viewed as a PT transformation from the point of view of *field* theory (Bell, 1955, *passim*); and Richard Feynman's slogan 'antiparticles are particles travelling backwards in time' is fairly naturally interpreted as suggesting the same idea. If this is correct, it *dissolves* the first puzzle: there *is* no such connection between Lorentz invariance and a genuinely 'non-spatio-temporal' symmetry.

The arguments in favour of interpreting the key transformation as CPT rather than as PT, however, can seem very compelling. One of the main aims of the paper of chapter 2 is to remedy this by exploring, in the context of *classical* electromagnetism, a 'time reversal' transformation that is the classical counterpart of the quantum transformation usually called 'TC', and arguing that it is no less plausible to regard this as the implementation of time reversal than it is to regard the transformation that is *usually* called 'time reversal' as the implementation of time reversal in classical electromagnetism. The project requires a thorough re-examination of the *meaning* of time reversal; the second main aim of this paper is to articulate and advocate a novel conception of time reversal (distinct from the usual notions of 'active' and 'passive' time reversal) that we think is implicit in recent work of David Malament (2004).

For the purposes of the remainder of the dissertation, I presuppose that the transformation usually called 'CPT' *is*, indeed, properly interpreted as a PT transformation.

The papers of chapters 3 and 4 also pursue the approach of attempting to illuminate the PT theorem by exploring the analogous issues in the classical context. Chapter 3 focusses on the issue of whether or not an analogue of the quantum-theoretic PT theorem can be proved in the context of classical field theory. The (surprising) answer is that it can for classical *tensor* field theories, but not for classical *spinor* field theories. The core of this paper is mathematical: it states and proves the classical PT theorem for tensor field theories.

Chapter 4 turns to our second puzzle: how it can be that a smaller symmetry (Lorentz invariance) entails a larger one, *even if* both are spacetime symmetries. In particular, I argue there that the existence of a PT theorem suggests that a

Lorentz-invariant field theory cannot make essential use of a representation of a preferred direction of time; thus our puzzle becomes the question of why *that* should be the case. After articulating this puzzle, I offer a solution based on the observation that there is no tensor that is Lorentz-invariant but not PT-invariant.

Chapter 2

Time reversal in classical electromagnetism

2.1 Introduction

> A backwards-moving electron when viewed with time moving forwards appears the same as an ordinary electron, except it's attracted to normal electrons — we say it has positive charge. For this reason it's called a 'positron'. The positron is a sister to the electron, and it is an example of an 'anti-particle'. This phenomenon is quite general. Every particle in Nature has an amplitude to move backwards in time, and therefore has an anti-particle. (Feynman, 1985, page 98)

Note that Feynman is not making any claims about backwards causation. He is merely claiming that if you time reverse a sequence of particle states you get a sequence of corresponding anti-particle states. According to standard quantum field theory textbooks this is not so: the *charge conjugation* operator turns particles into antiparticles, but time reversal does not. So we read Feynman as suggesting that the *real* time reversal operation (whatever that may mean — on which more below!) is not the operation that is usually given that name. Or, at least, that is the view that we are interested in comparing to the standard view, and that is the view we will call 'Feynman's view'.

Feynman's remarks, of course, were made in the context of quantum field theory. Meanwhile, in classical electromagnetism: David Albert (2000) has argued that classical electromagnetism is not time reversal invariant, because (according to him) there is no justification for flipping the sign of the magnetic field under time reversal. David Malament (Malament, 2004) has replied in defense of the standard view of time reversal, according to which the **B** field does flip sign and the theory is time reversal invariant.

Malament's discussion may leave one with the feeling that one only has to appreciate both (i) the four-dimensional formulation of classical electromagnetism and (ii) what we *mean*, or ought to mean, by 'time reversal', and the standard transformation $\mathbf{B} \stackrel{T}{\longmapsto} -\mathbf{B}$ will follow. This, however, is incorrect: there is an alternative to Malament's account, consistent with both (i) and (ii). It is an account according to which the magnetic field does *not* flip sign under time reversal (the *electric* field does), but the theory is time reversal invariant anyway; it is the classical analog of Feynman's view.

This paper has two main aims: (i) to explore the 'classical Feynman' view, with the hope that this may later illuminate important issues in quantum field theory, and, relatedly, (ii) to explore a novel conception of time reversal, distinct from the usual notions of 'active' and 'passive' time reversal, that we think is implicit in Malament's work and deserves further attention.

The structure of the paper is as follows. In section 2.2 we discuss the standard account of what time reversal is, and why one should care about it. Section 2.3 is a critical review of the existing debate concerning time reversal in classical electromagnetism: the standard 'textbook' account, Albert's objection, and Malament's reply. One of the things this discussion throws up is the contrast between Malament's notion of time reversal, which we call 'geometric' time reversal, on the one hand, and the familiar notions of 'active' and 'passive' time reversal on the other; in the course of discussing Malament's reply, we articulate the 'geometric' notion, and the interests relative to which it is the appropriate notion to focus on. In section 2.4 we articulate the 'Feynman' account, in terms of geometric time reversal. Section 2.5 investigates the possibility of 'deflating' the apparent dispute between the 'Malament' and 'Feynman' accounts, and regarding them as equivalent descriptions of the same underlying reality. Section 2.6 is the conclusion.

2.2 Time reversal and the direction of time

Let's start with the more-or-less standard account of what time reversal is, and why one should be interested in it.

Suppose we describe a world (or part of a world) using some set of coordinates x, y, z, t. A **passive time reversal** is what happens to this description when we describe the same world but instead use coordinates x, y, z, t', where $t' = -t$. An **active time reversal** is the following: keep using the same coordinates, but change the world in such a way that the description of the world in these coordinates changes exactly as it does in the corresponding passive time reversal. (So active and passive time reversal have exactly the same effect on the coordinate dependent descriptions of worlds.)

Suppose now that we have a theory which is stated in terms of coordinate dependent descriptions of the world, i.e. a theory which says that only certain coordinate dependent descriptions describe physically possible worlds. Such a theory is said to be **time reversal invariant** iff time reversal turns solutions into solutions and non-solutions into non-solutions. (Since active and passive time reversals have the same effect on the coordinate dependent descriptions of worlds, it follows that coordinate dependent theories will be invariant under active time reversal iff they are invariant under active time reversal.)

Why might one be interested in the time reversal invariance of theories? One reason (and the one we will be most interested in) is that failure of time reversal invariance of a theory indicates that time has an objective direction according to that theory. Why believe that? Well, suppose that we start with a coordinate dependent description of a world (or part of a world) which our theory allows. And suppose that after we do a passive coordinate transformation our theory says that the new (coordinate dependent) description of this world is no longer allowed. This seems odd: it's the same world after all, just described using one set of coordinates rather than another. How could the one be allowed by our theory and the other not? Indeed, this does not make much sense unless one supposes that the theory, as stated in coordinate dependent form, was true in the original coordinates but not in the new coordinates. And that means that according to the theory there is some objective difference between the x, y, z, t coordinates and the x, y, z, t' coordinates (where $t' = -t$). So time has an objective direction: that is, there is an objectively preferred temporal orientation. And if we want to write our theory in a coordinate independent way we are going to have to introduce a representation of this temporal orientation into our formalism.

Let's now clarify and modify this standard account a little bit. Let's start by asking a question that is rarely asked in physics texts, namely, what determines how things transform under a time reversal transformation? Well, space-time has some coordinate independent structure, and it is inhabited by coordinate independent quantities. We often describe that structure and those quantities in a coordinate dependent manner, but the structure of space-time itself is a coordinate independent geometric structure, and the quantities that inhabit space-time are coordinate independent quantities. This coordinate independent structure and those coordinate independent quantities determine what the coordinate dependent representations of that structure and of those quantities look like, and therefore determine how those coordinate dependent representations transform under space-time transformations. That's all there is to it.

Now, what we have just said might seem rather obvious, rather vague, and hence rather useless. However, there are a few important lessons to be learned from what we have said that are not always heeded.

Firstly, it means some quantities transform non-trivially (i.e. do not remain invariant) under time reversal. (Why it is worth noting this will become clear when we discuss David Albert's views on time reversal.)

Secondly, it means that it is not arbitrary how a quantity transforms under time reversal: how a quantity transforms under time reversal is determined by the (geometric) nature of the quantity in question, not by the absence or presence of a desire to make some theory time reversal invariant. For instance, one might think that one can show that some theory which, *prima facie*, is not time reversal invariant in fact is time reversal invariant, simply by making a judicious choice for how the fundamental quantities occurring in the theory transform under time reversal. However, if one changes one's view as to what the correct time reversal transformations are for the fundamental quantities occurring in a theory, then one is thereby changing one's view as to the geometric nature of those fundamental quantities, and hence one is producing a new, and different, theory of the world rather than showing that the original theory was time reversal invariant. That is to say, in such a circumstance one faces a choice: this theory with these quantities and these invariances or that theory with those quantities and those invariances. If the competing theories are empirically equivalent then one should make such a choice in the usual manner: on the basis of simplicity, naturalness, and so on.

Thirdly, even if a coordinate dependent formulation of a theory is not invariant under a passive time reversal, this does not yet imply that space-time must have an objective temporal orientation. For coordinate system x, y, z, t and coordinate system x, y, z, t' where $t' = -t$ not only differ in their temporal orientation, they also differ in their space-time handedness. So failure of invariance of the theory under time reversal need not be due to the existence of an objective temporal orientation, it could be due to the existence of an objective space-time handedness. That is to say, one might be able to form two rival coordinate independent theories, one of which postulates an objective temporal orientation but no space-time handedness, while the other postulates an objective space-time handedness but no temporal orientation. In order to decide which is the better theory, one will have to look at other features of the theories (such as other invariances).

More generally, what we want to know is what structure space-time has, and what quantities characterize the state of its contents. If we have in our possession an empirically adequate coordinate dependent theory, then what we should do is manufacture the best corresponding coordinate independent theory that we can, and see what space-time structure and what quantities this coordinate independent theory postulates. In fact, in the end the issue of what the correct time reversal transformation is is a bit of a red herring. What we are really interested in is what space-time structure there is and what quantities there are (and of course we are interested in the equations that govern their interactions). But the invariances and non-invariances of empirically adequate coordinate dependent formulations of theories are useful for figuring that out.

The above discussion was perhaps a bit abstract. So let us turn to a specific case which has been the subject of a fair amount of debate and controversy, namely that of classical electromagnetism.

2.3 Classical electromagnetism: the story so far

2.3.1 The standard textbook view

Let's start with the standard textbook account of time reversal in classical electromagnetism. The interaction between charged particles and the electromagnetic field is governed by Maxwell's equations and the Lorentz force law. In a particular coordinate system x, y, z, t, Maxwell's equations can be written as

$$\nabla \cdot \mathbf{E} = \rho \tag{2.1}$$

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j} \tag{2.2}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.3}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.4}$$

and the Lorentz force law can be written as:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{2.5}$$

Now let us ask how the quantities occurring in these equations transform under time reversal. According to the standard account the active time reverse of a particle that is moving from location A to location B is a particle that is moving from B to A. So, according to the standard view, the ordinary spatial velocity $\mathbf{v}$ must flip sign under active time reversal. Obviously, the current $\mathbf{j}$ will also flip over under active time reversal, while the charge density ρ will be invariant under time reversal.

Next let us consider the electric and magnetic fields. How do they transform under time reversal? Well, the standard procedure is simply to assume that classical electromagnetism is invariant under time reversal. From this assumption of time reversal invariance of the theory, plus the fact that $\mathbf{v}$ and $\mathbf{j}$ flip under time reversal while ρ is invariant, it is inferred that the electric field $\mathbf{E}$ is invariant under time reversal, while the magnetic field $\mathbf{B}$ flips sign under time reversal. Summing up, we have:

$$\mathbf{v} \stackrel{T}{\longmapsto} -\mathbf{v}; \tag{2.6}$$

$$\mathbf{j} \stackrel{T}{\longmapsto} -\mathbf{j}; \tag{2.7}$$

$$\mathbf{E} \stackrel{T}{\longmapsto} \mathbf{E}; \tag{2.8}$$

$$\mathbf{B} \stackrel{T}{\longmapsto} -\mathbf{B}; \tag{2.9}$$

$$\rho \stackrel{T}{\longmapsto} \rho; \tag{2.10}$$

$$\nabla \stackrel{T}{\longmapsto} \nabla; \tag{2.11}$$

$$t \stackrel{T}{\longmapsto} -t. \tag{2.12}$$

It follows from this time reversal transformation, as straightforward inspection of Maxwell's equations and the Lorentz force law can verify, that time reversal turns solutions into solutions and non-solutions into non-solutions.

2.3.2 Albert's proposal

David Albert ((2000, chapter 1)) takes issue with the textbooks' account of time reversal in classical electromagnetism. The point of contention is whether or not the magnetic field flips sign under time reversal. The standard account, we have seen, says that it does: $\mathbf{B} \xmapsto{T} -\mathbf{B}$. Albert suggests, however, that by 'time reversal' one ought to mean 'the *very same thing*' happening in the opposite temporal order; it follows (according to Albert) that the magnetic field (on a given timeslice) will be invariant under time reversal; and it follows from *that* (given Maxwell's equations) that the theory is not time reversal invariant. (Albert is happy with a non-trivial time reversal operation for, say, *velocity*. But that is because velocity is just temporal derivative of position, so of course *it* flips sign under time reversal. Albert's point is that the magnetic field is not the temporal derivative of anything.)

The difference in direction of argument between Albert and the textbooks is worth highlighting. In the textbooks' account reviewed above, the desideratum that the theory should be time reversal invariant enters as a *premise*. One finds some transformation on the set of instantaneous states that has the feature that, if it were the time reversal transformation, then the theory would be time reversal invariant, and one concludes that this is the time reversal operation. Albert is insisting on the opposite direction of argumentation: one should *first* work out which transformation on the set of instantaneous states implements the idea of 'the same thing happening backwards in time'; then and only then one should compare one's time reversal operation to the equations of motion, and find out whether or not the theory is time reversal invariant. He is further insisting that, in the case of electromagnetism, this has not adequately been done.

Albert has a point here. One should, indeed, be wary of taking the textbooks' strategy to extremes: it is not difficult to show that, under very general conditions, *any* theory, including ones that are (intuitively!) not time reversal invariant, can be made to come out 'time reversal invariant' if we place *no* constraints on what counts as the 'time reversal operation' on instantaneous states. This problem for an unconstrained version of the textbook strategy is the 'triviality problem'. In the next section, we set this problem out in more detail.

2.3.3 The triviality problem

Consider an arbitrary physical theory that furnishes a set S of *instantaneous states*. We can then form the set H_{kin} of *kinematically allowed histories* as follows: H_{kin} is the set of all functions h assigning an instantaneous state $h(t) \in S$ to each time $t \in \mathbb{R}$. The theory will tell us that some, but not all, of these histories are *dynamically* allowed. Thus, we have a set H_{dyn} of dynamically allowed histories; $H_{dyn} \subset H_{kin}$.

Let the notions of *past-* and *future-determinism* be defined as follows. Say that a theory is *past-deterministic* iff the state at any given time determines the state at all earlier times, in the following sense:[1]

Past-determinism. $\forall h_1, h_2 \in H_{dyn}, \forall t_1, t_2 \in \mathbb{R}, (h_1(t_1) = h_2(t_2)) \rightarrow$
$(\forall t' < t_1)\,(h_1(t_1') = h_2(t_1' + (t_2 - t_1)))$.

Similarly, say that a theory is *future-deterministic* iff the state at any given time determines the state at all later times:

[1]This condition builds in a requirement of time translation invariance, over and above that of determinism as the latter is usually understood. The stronger condition is required for the argument below.

Future-determinism. $\forall h_1, h_2 \in H_{dyn}, \forall t_1, t_2 \in \mathbb{R}, (h_1(t_1) = h_2(t_2)) \rightarrow$
$(\forall t' > t_1)\,(h_1(t_1') = h_2(t_1' + (t_2 - t_1)))$.

These notions can be combined into a single, temporally symmetric notion of determinism: Say that a given theory is *deterministic* iff it is both future-deterministic and past-deterministic. We will confine our attention to theories that are deterministic in this sense.

The triviality problem is the following: if one allows completely arbitrary time reversal operations, then *any* theory that is both future- and past-deterministic will count as time reversal invariant.

Here is the argument for this claim. Take an arbitrary theory that has a formulation with the structure sketched above: that is, let the theory specify a set S of instantaneous states and a set H_{dyn} of dynamically allowed histories. We construct a 'time reversal operation' on S in the following way:

- Note that the assumptions of past-and future-determinism guarantee that the following relation on the space H_{dyn} of dynamically allowed histories is an equivalence relation:

$$h_1 \sim h_2 \text{ iff } (\exists t)(h_2(t) = h_1(0)). \tag{2.13}$$

- Choose a representative state from each equivalence class; let $\tilde{H}_{dyn} \subset H_{dyn}$ denote the set of representative histories.

- Now, for any state $s \in S$, there will be a unique history $h_s \in \tilde{H}_{dyn}$ and a unique time t_s such that $h_s(t_s) = s$.

- Define the 'time-reversal' operation $R : S \rightarrow S$ as follows:

$$\forall s \in S, R(s) = h_s(-t_s). \tag{2.14}$$

This construction guarantees that, if $h \in H_{dyn}$ is any dynamically allowed history, then so also is $\overline{R}(h)$, where the latter is defined as follows:

$$\forall h \in H_{dyn}, \forall t \in \mathbb{R}, \left(\overline{R}(h)\right)(t) = R(h(-t)). \tag{2.15}$$

That is, the theory is guaranteed to be 'time reversal invariant' relative to the 'time reversal operation' R. If we were pushing the textbooks' line of argument to its logical conclusion, therefore, we would conclude that this operation R did indeed count as implementing time reversal.

Why is this a *problem*? Well, the point is that one can very easily cook up theories that are past- and future-deterministic, but that are obviously *not* 'time reversal invariant' in any remotely intuitive or useful sense. Therefore, the fact that the textbook strategy will always succeed in rendering such a theory 'time reversal invariant' ought to raise a good deal of suspicion about the validity of that strategy.

Here is a toy example: an example, that is, of a theory that is past- and future-deterministic, but that (intuitively!) is not time reversal invariant. The instantaneous state space is $R^3 - 0$: the instantaneous state gives the location of a single point particle in a three-dimensional space at the time in question, and there is a privileged point O (the origin) that the particle is forbidden from occupying. The dynamics is as follows: the particle always moves towards O, with a speed proportional to its distance from O. It is easy to see that this theory is past- and future-deterministic. It is equally easy to see that the theory is not time-reversal invariant in any reasonable sense of 'time reversal'. However, if we were to push the textbooks' line of argument to its logical conclusion, we would find ourselves saying: the theory is time reversal invariant, it's just that the time reverse of a particle at position p is a particle at some other position $q \neq p$.

This shows that the textbooks' direction of argument is in danger of begging the question regarding the time reversal invariance of a given theory. Some theories are *not* time-reversal invariant, so we had better not make a fully general practice of assuming, as a premise in a derivation of the time-reversal operation on instantaneous states, that the theory *will* be time-reversal invariant.

The conclusion to draw is that, if discussing the time reversal invariance of theories is to be a game worth playing, there must be *some* constraints on which transformations on the set of states one is allowed to call 'time reversal operations'. One cannot just write down any old transformation and call it 'time reversal'; one must explain why the operation in question deserves the name.

This raises the question of what exactly the constraints should be — what, that is, it should take for some given verbal performance to count as an 'explanation' or 'justification' of a time reversal operation. The answer to this is interest-dependent: it depends on why one is interested in the question of time-reversal invariance in the first place. We are aware of two sorts of reasons for caring about the time-reversal invariance of theories — those we will label by the names 'the pragmatic program' and 'the ontological program' — and, correspondingly, two well-motivated responses to the triviality problem. Spelling these out is the task of sections 2.3.4 and 2.3.5.

2.3.4 Justifying time reversal operations: the pragmatic program

The first notion of 'justification' is relatively liberal. The constraint is merely that the operation in question must capture the notion of 'the same thing happening backwards in time' in some reasonably intuitive sense. For example, if a given history describes a baseball moving to the left, then any history that describes a baseball moving to the right (at the same speeds, etc.) will be well-qualified to count as the 'time reversal' of the first. This may yet leave the time reversal operation indeterminate, but no matter: any candidate satisfying these minimal, intuitive constraints will do, and there need be no fact of the matter as to which is the correct one. Call any such transformation a 'pseudo time reversal' operation.

There are good reasons to be interested in the pseudo time reversal invariance of theories. Perhaps the most obvious is the much-discussed issue of the emergence of macroscopic time-asymmetry from 'time reversal invariant' microphysics. It does not matter, for the purposes of framing and trying to answer the question of why ripples often spread outwards on a pond but never converge to a point, why eggs often break but never spontaneously mend, etc, whether the microphysics that generates the puzzle is merely 'pseudo time reversal invariant', or time reversal invariant in any more elevated sense. (Even Albert agrees with this: while he insists that the operation that sends $\mathbf{E}$ to itself and $\mathbf{B}$ to $-\mathbf{B}$ is not really time reversal, he agrees that the invariance of the theory under this operation is quite enough to generate a puzzle about, say, the asymmetry of radiation.)

A similar attitude to time reversal is taken by Robert Geroch, in the context of quantum field theory. If one has found more than one automorphism of the operator algebra that seems to match the intuitive idea of time reversal reasonably well, and if the theory is invariant under both, then, says Geroch, there will be no question about which is 'really' the time reversal operation:

> The point is that *any* operation that commutes with the S-matrix is valuable. We regard the words ['time reversal'][2] as merely suggesting a particularly fertile area in which such operators might be found. This philosophy is important . . . (Geroch, 1973, page 104)

Note that, if one's research program is the pragmatic program, then the form of argument that we earlier accused of 'question-begging' is perfectly acceptable. There is a puzzle about the emergence of macroscopic time-asymmetry from an

[2]Geroch is actually discussing charge reversal, rather than time reversal, in the passage quoted, but he states that he takes the same attitude to time reversal (*ibid.*, p.107).

underlying microphysics that is time reversal invariant in *any* sense of time reversal that is permitted by the pragmatic program; the quantum field theorist will have theoretical reasons to be interested in *any* operation that commutes with the S-matrix. So, if this is where one's interests lie, it is only rational to search for transformations under which the theory is invariant, and give names (like 'time reversal') to the transformations one finds.

2.3.5 Justifying time reversal operations: the ontological program

The pragmatic program is willing to let many flowers bloom. But there is also a more demanding view of time reversal, according to which only one flower may bloom: that is, if one operation represents time reversal in this more elevated sense, then no other operation can be an equally good deserver of the name. This more demanding view of time reversal is associated with a different research program, which we call the 'ontological program'. This will be the program we are interested in pursuing for the remainder of the paper.

The key idea of the ontological program is the one we have already stated, in section 2.2: the time reversal invariance or non-invariance of our best physical theories is intimately related to the question of whether or not time has an objective direction — the question, that is, of whether or not a distinguished temporal orientation is part of the spatiotemporal furniture of the world. In this section, we explain this idea in more detail.

Spacetime structure from time-reversal invariance: a first pass.

Let us take up an abstract, coordinate-independent point of view. Suppose we are dealing with a theory whose ontology, we currently think, contains a spacetime manifold M, a set of (other) geometric objects Ψ on M, and a temporal orientation

τ. A *model* of the theory is then given by a triple $\langle M, \Psi, \tau\rangle$. However, not all such triples will be models of the theory; the theory will specify some constraints on how the various fields are to be related to one another, and (perhaps) to the temporal orientation.

We then ask ourselves the following question: is the class of models invariant under the transformation

$$\langle M, \Psi, \tau\rangle \mapsto \langle M, \Psi, -\tau\rangle? \tag{2.16}$$

If the answer is 'no', then we must take the temporal orientation τ seriously: since there is some pair of triples, one of which is permitted by the theory and the other of which is not, but which differ from one another only on the direction of the temporal orientation, it must be that changing the temporal orientation amounts to making an objective change to the world. So the temporal orientation must be physically real. If the answer is 'yes', on the other hand, then good methodological practice (Ockham's Razor) urges us to excise the temporal orientation τ from the ontology of the theory.

Time-reversal invariance of coordinate-dependent formulations.

For the purposes of discussing coordinate-dependent formulations, let us make the simplifying assumption that our spacetime manifold M is diffeomorphically equivalent to $\mathbb{R}^4$. Then, M can be covered by a single coordinate chart $\phi : M \to \mathbb{R}^4$.

For a fixed triple $\langle M, \Psi, \tau\rangle$, relative to a given such chart, we will then have a coordinate-dependent representation of each of our dynamical objects $\psi \in \Psi$ as a field (or similar) on $\mathbb{R}^4$. The theory's constraints are then represented as equations constraining the coordinate-dependent representations of the dynamical objects.

Now, suppose that we do not explicitly take a coordinate-dependent representation of the temporal orientation τ, but rather we 'represent' τ in the following way: we agree only to use coordinate charts with the feature that the time coordinate t increases in the direction picked out by the orientation τ. That is, in discussing a coordinate-dependent representation of the dynamical objects Φ in some particular chart ϕ whose time coordinate increases in the temporal direction picked out by τ, we take ourselves to be discussing the model $\langle M, \Phi, \tau \rangle$ rather than $\langle M, \Phi, -\tau \rangle$, where τ is the temporal orientation that agrees with t.

If we engage in this practice, we are not using a generally covariant formulation of the theory. There is then the possibility that, while the coordinate-dependent constraining equations correctly represent the theory in the coordinate chart $\phi : M \to \mathbb{R}^4$ for which they were derived, *those same equations* would not correctly represent the theory relative to the 'time-reversed' coordinate chart $\phi' : M \to \mathbb{R}^4$, where $\phi^{-1}(t, x, y, z) = \phi'^{-1}(-t, x, y, z)$. This failure of covariance would be the coordinate-dependent symptom of the fact that the theory has some models $\langle M, \Phi, \tau \rangle$ for which the 'time-reverse' $\langle M, \Phi, -\tau \rangle$ is not a model.

At this point, one might well wonder what the point is of using coordinates at all, if one's interest is in the structure of spacetime (as opposed to, say, calculational convenience). After all, the process we have just described — of reconstructing the coordinate-dependent formulation of a theory from its coordinate-independent counterpart in a way that (*if* the class of models fails to be invariant under $\langle M, \Phi, \tau \rangle \mapsto \langle M, \Phi, -\tau \rangle$) will fail to be generally covariant, and then testing the time-reversal covariance of the resulting coordinate-dependent equations — would seem to be a rather inefficient way of getting at the question of the time reversal invariance or non-invariance of the theory. Surely a simpler way to find out whether or not the theory is time reversal invariant would be to eschew talk of coordinates altogether, and directly to test the invariance of the class of models under the transformation $\langle M, \Phi, \tau\rangle \mapsto \langle M, \Phi, -\tau\rangle$?

Well, yes. In particular, if one is lucky enough already to have an intrinsic formulation of one's theory, then, purely for the purposes of identifying the pieces of spacetime structure to which one should take the theory to be ontologically committed, there is no point whatsoever in passing to a coordinate-dependent formulation. There are two reasons, however, for being interested in the shape taken by the time-reversal invariance issue in a coordinate-dependent formulation of a theory. The first is that, as a matter of fact, one often does not (yet) *have* an intrinsic formulation of one's theory, and one wants to know what one should conclude about the structure of spacetime from the covariance and non-covariance properties of one's (coordinate-dependent) dynamical equations under various coordinate transformations. The above sketch of the relationship between time reversal in an intrinsic formulation of a theory and time reversal in a coordinate-dependent formulation is useful, in the first instance, for understanding what it is we are asking (and why) when we ask questions about the covariance properties of coordinate-dependent equations. (In particular, this understanding will dictate what it is to justify a time-reversal operation on coordinate-dependent descriptions.) The second reason is still more practical: we wish to make contact with discussions of time reversal in physics textbooks, and such discussions are usually carried out in a coordinate-dependent language.

To sum up: Our more demanding notion of 'justification' for a candidate time-reversal operation on coordinate-dependent, instantaneous states is motivated by the following: the operation must be such that we can legitimately draw conclusions about the structure of spacetime, based on the covariance or non-covariance of the theory's coordinate-dependent equations under the time reversal operation based on *this* transformation on S. The 'pragmatic' notion of justification is too liberal for this purpose. Below, we will offer a notion that is perfectly suited to it.

So something in Albert's objection seems to be right: justification of non-trivial time reversal operations is required. We do not, however, endorse his account of time reversal in electromagnetism. We will come back to this after discussing an alternative account, due to David Malament.

2.3.6 Malament's proposal

Malament seeks to justify the usual textbook time reversal operation for classical electromagnetism, and for the **B** field in particular.

At first sight, one might think that this is done as soon as one thinks relativistically, and conceives of the **E** and **B** fields as components of the Maxwell-Faraday tensor F^{ab}. A moment's thought, however, shows that this is not the case. The electric field is read off from the space-time components of F^{ab}, while the magnetic field is read off from the space-space components:[3]

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_1 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}. \tag{2.17}$$

If the Maxwell-Faraday tensor F^{ab} itself (as a tensor) is invariant under time reversal, then it will be the *electric* field, not the magnetic field, that flips sign when we perform a passive time reversal. To justify the standard textbook transformation, we need to justify a sign flip for F^{ab}: $F^{ab} \xmapsto{T} -F^{ab}$. This is the task that Malament takes up.

[3]Roman subscripts and superscripts indicate that we are using the abstract index notation: F^{ab} is a rank two tensor, not a component of such a tensor in a particular coordinate system. When we wish to refer to coordinate-dependent components of tensors, we use Greek indices, as in $F^{\mu\nu}$.

Malament's treatment of electromagnetism embodies a particular conception of what it *means* to 'justify' a time reversal operation, and, relatedly, a third conception (alongside active and passive time reversal) of what time reversal *is*. We will first state these explicitly (but somewhat abstractly), then let our exposition of Malament's treatment of electromagnetism illustrate them:

- To give a **justification** of a non-trivial time reversal operation $X \stackrel{T}{\longmapsto} X'$ for a state description X is to postulate a particular fundamental ontology for the theory, and to explain how the representation relation between X and the objects of the fundamental ontology depends on temporal orientation, in such a way that it follows that if we flip the temporal orientation but hold the remainder of the fundamental objects fixed, the state description changes as $X \stackrel{T}{\longmapsto} X'$.

- **Geometric time reversal:** To time-reverse a kinematically possible world, hold all the fundamental quantities fixed [with the exception of the temporal orientation, if that is a fundamental object], and flip the temporal orientation.

These notions of justification and of time reversal are, of course, the ones that are perfectly suited to the 'ontological program' discussed in section 2.3.5.

Malament's treatment of electromagnetism. Malament's account is as follows. There are two fundamental types of objects in a classical electromagnetic world. There are charged particles, and there is the electromagnetic field. Now, the dynamics happens to be such that it will be convenient, mathematically, to represent the motions of particles by means of four-velocities, where the four-velocity at any point on the worldline is tangent to the worldline at that point. The crucial fact now is that a world-line does not have a unique tangent vector at a point: at each point on a world-line, there is a continuous infinity of four-vectors that are tangent to the world-line at the point in question. We can narrow

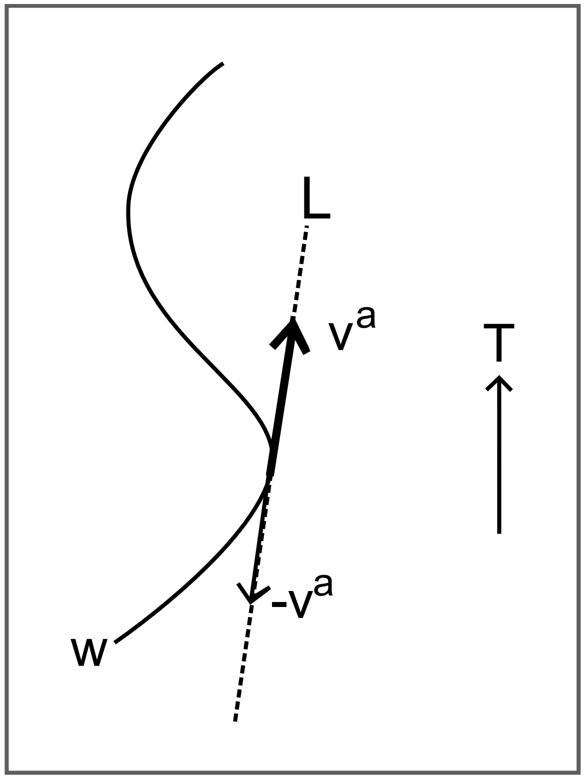


Figure 2.1: w is the worldline of a particle of mass m and charge q. L is the tangent line to w at the point p. Until we have specified a temporal orientation τ, we have left it open whether the four-velocity is v^a or $-v^a$. $v^b\nabla_b v^a$ is the four-acceleration; it is independent of temporal orientation. The electromagnetic field F maps $< L, q >$ to the four-force $mv^b\nabla_b v^a$.

things down somewhat by stipulating that four-velocities are to have unit length, but this still does not quite do the trick: one can associate *two* unit-length four-vectors that are tangent to the world-line at the point in question (if $v^a \in T_p$ is one, then $-v^a$ is the other; see figure 2.1).

Next, how should we conceive of an electromagnetic field at a point p in spacetime? According to Malament, we should think of the electromagnetic field at p as a quantity which, for any tangent line L at p and charge q, determines what 4-force a (test) particle with charge q and tangent line L at p would experience. More formally, Malament conceives of the electromagnetic field F (*not* F^{ab}) at a point p as a map from pairs $\langle L, q \rangle$ at p to four-vectors at p.

How do Malament's fundamental quantities (tangent lines, maps from tangent lines to 4-vectors) relate to the standard quantities (4-vectors, Maxwell-Faraday tensor) occurring in our three equations? The relation is simple: *relative to a choice of temporal orientation*, one *can* associate a unique unit-length tangent vector with each location on a timelike world-line, namely, the one that is 'future'-directed according to that temporal orientation. So, given a temporal orientation, we can represent any given tangent line by a unique unit length four-vector, i.e. a four-velocity. Given such a representation, the electromagnetic field can be represented by a linear map from four-vectors to four-vectors. And that just means that, given a temporal orientation we can represent the electromagnetic field as a rank 2 tensor, which we can identify as the standard representation of the electromagnetic field by the Maxwell-Faraday tensor F^{ab}.

So, given a temporal orientation, Malament can formulate classical electromagnetism using the usual covariantly-formulated equations: the Maxwell equations,

$$\nabla_{[a} F_{bc]} = 0, \tag{2.18}$$

$$\nabla_n F^{na} = J^a, \tag{2.19}$$

and the Lorentz force law,

$$qF^a{}_b V^b = m v^b \nabla{}_b v^a. \tag{2.20}$$

Using the geometric conception of time reversal, it is then straightforward to see how the quantities in these equations transform under time reversal. Recall that on the geometric conception, to 'time reverse' is to leave all the *fundamental* quantities fixed, and to flip temporal orientation. We then hold fixed (also) our conventions about how non-fundamental quantities are derived from the fundamental ones in an orientation-relative way, and we see which transformations for the non-fundamental quantities result. Now, on Malament's picture, four-velocity is not fundamental: it is defined only relative to a choice of temporal orientation. If v^a is the four-velocity, i.e. is the unit-length future-directed tangent to a given worldline at some point p relative to our original choice of temporal orientation, then $-v^a$ will be the four-velocity relative to the opposite choice of temporal orientation. Similarly, if F^{ab} correctly maps four-velocities to four-forces relative to our original orientation, then, in order to represent *the same map from tangent lines to four-forces* relative to the opposite choice of temporal orientation, we will have to flip the sign of the tensor, to compensate for the sign flip in four-velocity: $F^{ab} \mapsto -F^{ab}$. We have now given justifications for Malament's time reversal operations for v^a and F^{ab}:

$$v^a \xmapsto{T} -v^a; \tag{2.21}$$

$$F^{ab} \xmapsto{T} -F^{ab}. \tag{2.22}$$

Electric and magnetic fields.

As Malament notes, the frame-independent formulation suffices to write down the dynamics of the theory and establish their time-reversal invariance. Like Malament, however, we wish to make contact with Albert and the textbooks; to do this, we need to consider decompositions of our four-dimensional F^{ab} into electric and magnetic fields.

Following Malament ((2004, pages 16–17)), we make the following two definitions:

- A **volume element** ϵ_{abcd} on M is a completely antisymmetric tensor field satisfying the normalization condition $\epsilon_{abcd}\epsilon^{abcd} = -24$.

- A **frame** η_a is a future-directed, unit, timelike vector field that is constant ($\nabla_a \eta^b = 0$).

We can now decompose the electromagnetic field into electric and magnetic fields, relative to a given frame and volume element:

$$E^a := F^a{}_b \eta^b; \tag{2.23}$$
$$B^a := \frac{1}{2}\epsilon^{abcd}\eta_b F_{cd}. \tag{2.24}$$

Note that the electric field E^a and is defined relative to temporal orientation and frame; the magnetic field B^a is defined relative to temporal orientation, frame and volume element. The volume element itself is a more subtle case; we follow Malament in stipulating that it, too, flips sign under time reversal.[4]

It follows that (as Malament explains) the time reversal transformation acts as follows:

$$\tau \xmapsto{T} -\tau; \tag{2.25}$$
$$\eta_a \xmapsto{T} -\eta_a; \tag{2.26}$$
$$\epsilon_{abcd} \xmapsto{T} -\epsilon_{abcd}; \tag{2.27}$$
$$\upsilon^a \xmapsto{T} -\upsilon^a; \tag{2.28}$$
$$F^{ab} \xmapsto{T} -F^{ab}; \tag{2.29}$$
$$E^a \xmapsto{T} E^a; \tag{2.30}$$
$$B^a \xmapsto{T} -B^a. \tag{2.31}$$

[4]The point here is just that we choose to mean, by ‘time reversal’, ‘flip the temporal orientation and hold the spatial handedness fixed’ (so the total orientation, represented by the sign of the volume element, has to flip), rather than ‘flip the temporal orientation and hold the total orientation fixed’ (in which case the spatial handedness would have to flip).

Note that the electric field, E^a, is invariant under time reversal, while the magnetic field, B^a, flips sign. This is exactly the time reversal operation suggested by standard textbooks in classical electromagnetism. So, Malament's proposal provides a justification, based on his geometric conception of time reversal, for the standard view.

2.3.7 Albert revisited

We noted that, as soon as one thinks of the **E** and **B** fields as derived from a more fundamental Maxwell-Faraday tensor, either **E** or **B** must flip sign under time reversal. On Albert's account, neither flips sign. But, of course, Albert is perfectly aware of the four-dimensional formulation of electromagnetism. So why does he say what he says?

Well, on Albert's view, *pace* any arguments for interpreting electromagnetism in terms of a Minkowski spacetime, spacetime is in fact Newtonian, velocities are good old spatial 3-vectors, and so are the electric and magnetic fields.[5] The dynamics governing the development of the **E** and **B** fields, and the particle worldlines, happens to be 'pseudo-Lorentz invariant': that is, there exist simple transformations on the **E** and **B** fields such that, *if* those were the ways **E** and **B** transformed under Lorentz transformations, *then* the theory would be Lorentz invariant. This is perhaps surprising — there's no *a priori* reason to expect the dynamics to have this feature of 'pseudo Lorentz invariance', if one thinks that spacetime is Newtonian. But then, there's no *a priori* reason why the dynamics in a Newtonian spacetime *shouldn't* be pseudo Lorentz invariant, either. Similarly: it follows from this pseudo Lorentz invariance that observers will never be able to discover, merely by means of 'mechanical experiments' (i.e. observations of

[5]To our knowledge, Albert has not stated this view in print. Our attribution of it to him is based on conversations between Albert and one of us over a period of several years. We also do not know whether he still holds the view in question.

particle worldlines), what their absolute velocity is, or pin down the **E** and **B** fields uniquely. So *if* one thought that all features of reality must be empirically accessible to the human machine with its coarse-grained perceptive capacities, one would be very suspicious of Albert's view; but why, Albert might well ask in reply, should one think that?

What should one make of all this? Well, while we agree that Albert's view is internally coherent, we regard it as insufficiently motivated, for the following reason. A straightforward application of Ockham's razor prescribes that, faced with a choice between two empirically equivalent theories, one of which is strictly more parsimonious than the other as far as spacetime structure goes, one should (*ceteris paribus*) prefer the more parsimonious theory. In other words, one should commit to the *minimum* amount of spacetime structure needed to account for the empirical success of one's theories. Now, on Albert's view, spacetime is equipped with a preferred foliation and a standard of absolute rest; further, it must also be equipped with an objective temporal orientation, in order to account for the non-time-reversal invariance of classical electromagnetism. On the Minkowskian view, spacetime has none of this structure. If other things are equal, this gives us a reason to prefer a Minkowskian view; further, as far as we can see, other things *are* equal. We conclude that, insofar as classical electromagnetism is to be trusted at all, spacetime is Minkowskian rather than Newtonian, it is the unified electromagnetic field, rather than the **E** and **B** fields separately, that is fundamental, and that Albert's view of time reversal is false.

We will say no more about Newtonian interpretations. What is more interesting, for the purposes of our paper, is that even *given* a Minkowskian interpretation of relativity, the ontology, and hence the time reversal operation, for classical electromagnetism remains underdetermined. Malament has suggested one candidate ontology; we turn now to alternatives.

2.4 The 'Feynman' proposal

In this section, we turn to the view of time reversal that will correspond to Feynman's view of antiparticles. Our discussion here will not differ from our discussion of Albert's or Malament's proposals in terms of what time reversal is or how non-trivial time reversal operations are justified; that is, we are still thinking in terms of geometric time reversal. The 'Feynman' proposal is simply a different proposed ontology, a different view as to what fundamental quantities there in fact are out there in nature. It provides an geometric justification for a *third* time reversal operation for the electric and magnetic fields, distinct from both Albert's and Malament's.

Fundamental ontology.

The distinctive feature of the 'Feynman' proposal is the suggestion that there is a fundamental, temporal orientation-independent fact as to the sign of the four-velocity of a given particle. That is, we change our hypothesis about the fundamental properties possessed by particles: rather than supposing that particles' worldlines are mere sets of spacetime points, and hence intrinsically *undirected*, we now suppose that particles' worldlines are intrinsically directed: each worldline comes equipped with an arrow, and there is an objective, temporal-orientation-independent fact about which way the arrow on any given worldline points. In that case, we no longer have Malament's motivation for saying that the electromagnetic field is a map from *tangent lines* to four-vectors. So, on the 'Feynman' proposal, we take the electromagnetic field to be (fundamentally!) a map from *four-vectors* to four-vectors, or, equivalently, a rank 2 tensor field. Thus, the electromagnetic field, independent of a temporal orientation, corresponds to a unique rank 2 tensor: the Maxwell-Faraday tensor F^{ab}.

The **electric and magnetic fields**, E^a and B^a, are then defined from F^{ab}, relative to a frame and volume element, just as they are on Malament's proposal.

Time reversal.

The corresponding time reversal transformation is:

$$\tau \overset{T}{\longmapsto} -\tau \tag{2.32}$$

$$\epsilon_{abcd} \overset{T}{\longmapsto} -\epsilon_{abcd} \tag{2.33}$$

$$\eta^a \overset{T}{\longmapsto} -\eta^a \tag{2.34}$$

$$F^{ab} \overset{T}{\longmapsto} F^{ab} \tag{2.35}$$

$$v^a \overset{T}{\longmapsto} v^a \tag{2.36}$$

$$E^a \overset{T}{\longmapsto} -E^a \tag{2.37}$$

$$B^a \overset{T}{\longmapsto} B^a. \tag{2.38}$$

Note that this is not the textbook time-reversal transformation. The Feynman proposal has the consequence that the *electric* field flips sign under time reversal, and that the magnetic field does not — but it, too, has the consequence that the theory is time reversal invariant.[6]

More on the 'Feynman' proposal.

Certain features of the time-reversal operation sanctioned by the 'Feynman' proposal seem rather odd, however; let's take a closer look. Consider, for example, a particle travelling between Harry and Mary (see figure 2.2). Suppose that, prior

[6]The time reversal invariance of this theory is easy to see, by looking at the Lagrangian $L = -\frac{1}{4}F_{ab}F^{ab} - qv_aA^a$. Under 'Feynman' time reversal, all four of the objects appearing in this Lagrangian — the Maxwell-Faraday tensor F^{ab}, the charge q, the four-velocity v^a and the four-potential A^a — are invariant under time reversal. So of course the Lagrangian itself (a scalar field on M) is invariant under time reversal, and, consequently, there will never be a set of field configurations and particle worldlines that is dynamically permitted relative to one temporal orientation and not the other.

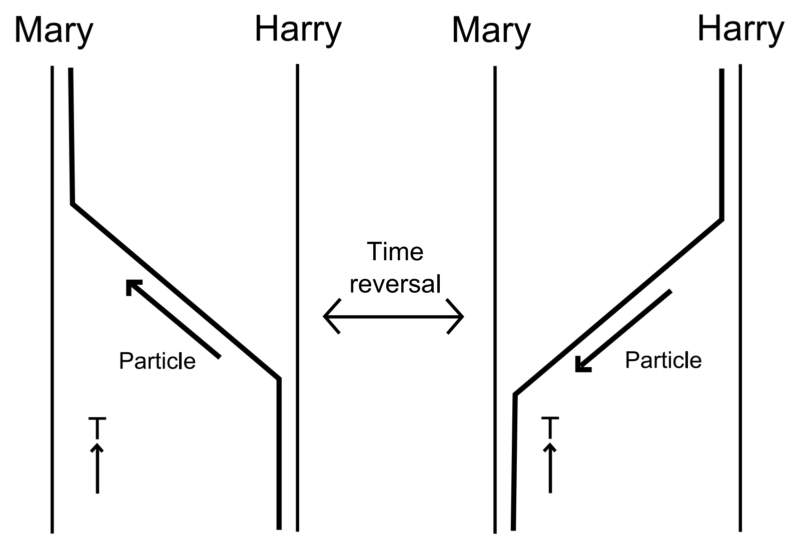


Figure 2.2: The time-reverse of a particle traveling from Mary to Harry, according to the Feynman view, is (still) a particle traveling from Mary to Harry.

to time reversal, the particle's four-velocity happens to be 'future'-directed, and points from Harry's worldline to Mary's. Then, the following two observations can be made about the time-reversed situation. First, in the time-reversed situation the particle's four-velocity will be 'past'-directed. (This follows from the fact that the four-velocity itself dows not change, while the description of a given temporal direction on the manifold as 'future'/'past' does change when we flip the temporal orientation.) Second, the four-velocity *will still point from Harry to Mary*. On the 'Feynman' proposal, that is, we are asked to make sense of a notion of 'time reversal' according to which the time-reverse of a particle traveling from Harry to Mary is not a particle traveling from Mary to Harry. This seems an odd feature of the 'Feynman' view.

However, let us suppose that it is not the case that the four-velocities of all particles point in the same temporal direction. That is, let us suppose that, relative to a fixed choice of temporal orientation, some particles have future-directed four-velocities, and others have past-directed four-velocities. Suppose, then, that we have a model of electromagnetism which consists of a single particle of charge q, moving in an electromagnetic field F^{ab} with four-velocity v^a. One can then trivially produce another model by keeping the electromagnetic field F^{ab} the same and the trajectory the same, while flipping the sign of the charge ($q \mapsto -q$) and of the four-velocity ($v^a \mapsto -v^a$). (One can see that this operation does indeed turn models into models by inspecting Maxwell's equations and the Lorentz force law, or, alternatively, by inspecting the Lagrangian. The only changes in any of these quantities are in the signs of q and v^a, which always occur together, so that the changes cancel; so, changing the sign of the charge and of the four-velocity must turn a solution into a solution, and a non-solution into a non-solution.)

Let us put this another way: a particle with charge q and four-velocity v^a behaves, in a given electromagnetic field, exactly as if it is a particle with charge $-q$ and velocity $-v^a$: it follows exactly the same trajectory, so that, given only access to the results of 'mechanical experiments', the two possible situations cannot be distinguished in any way. This observation opens the door for the following hypothesis: particles that we have regarded as belonging to different types, related by the 'is the antiparticle of' relation — electrons and positrons, say — are really of the same type as one another. In particular, they have the same electric charge as one another. Things appear otherwise only if we erroneously assume that all four-velocities must point in the same temporal direction as one another. In other words, we can achieve parsimony in particle types at the cost of the 'extravagance' of endowing particle worldlines with an intrinsic direction; the Feynman proposal is that we do so. If this hypothesis is right, then it is indeed true that an anti-particle is nothing but a particle traveling in the opposite direction of time.

2.5 Structuralism: A Third Way?

We have been assuming so far that the Malament and Feynman proposals represent distinct alternatives, at most one of which can be correct. One can have a different time reversal operation for the same formalism, we said, only if one makes a different postulate about the *fundamental ontology*; but if one does that, then (we said) one has changed one's theory, in the clear sense that one has changed one's hypothesis about the fundamental nature of the world.

Be that as it may, one might still (on the other hand) have the gut feeling that the 'disagreement' between the Malament and Feynman ontologies is not a genuine one; that the two 'rival theories' are, in some sense, saying the same thing in different ways.

Clearly, one cannot fully hold onto both of these ideas: one says that the Malament and Feynman proposals are distinct, the other says they are not. In the present section, however, we will sketch a third set of hypotheses about the fundamental nature of a classical electromagnetic world that does justice to the basic principles behind both ideas. It will do justice to the just-mentioned gut feeling, in that it will provide a way of regarding the claim that worldlines have arrows on them and that four-velocities can be past-directed (as Feynman says), and the claim that worldlines have no intrinsic arrows and four-velocities are always future directed (as Malament says), as equivalent descriptions of the same underlying situation. However, it will also do justice to our earlier insistence that this business of formulating alternative descriptions is not ontologically innocent, because it will be a third, rival, suggestion for what the fundamental nature of electromagnetic reality might be, rather than a claim that the *original* Malament and Feynman theories are equivalent.

The 'third way' is *structuralism*. In the broader context, structuralism arises as an attempt to steer the correct course between (on the one hand) an excessively deflationary positivism, according to which empirical equivalence is supposed straightforwardly to entail equivalence of meaning, and (on the other) an excessively realistic position, according to which every difference in notation (the use of the boldface letter $\mathbf{D}$ rather than $\mathbf{E}$ for the electric field, say) is taken to correspond to a difference in postulated physical reality. The sort of 'structuralism' we are interested in typically proceeds – either on a case-by-case basis (i.e. applying the structuralist strategy where and only where it happens to seem appropriate) or as a sweeping claim about the possibility of knowledge, reference and/or the nature of reality – by reifying, at the *fundamental* level, *relations*, but not monadic properties. This (fundamental reification of relations only) will be our tactic here too.

2.5.1 Structures: the debate recast

Before setting out the relationist's attempt to deflate the debate between the Malament and Feynman views, it will serve the interests of clarity if we recast the moves that have been made so far in a more formal framework.

In the beginning , we were representing classical electromagnetic worlds using one-parameter families of *standard Newtonian structures.* A standard Newtonian structure is a mathematical entity of the form

$$\mathcal{S}_{Newt} = \langle \Sigma \times T, P, \mathbf{x}, m, q_s, \mathbf{E}, \mathbf{B} \rangle, \tag{2.39}$$

where:

- $\Sigma \times T$ is a Newtonian spacetime: that is, Σ is a Euclidean three-space, and $T \sim (\mathbb{R}, +)$ is the set of times.
- P is a set of particles. (In the first instance, P is structureless; structure is

added by the functions $\mathbf{x}, m, q_s$ below.)

- $\mathbf{x} : P \times T \to \Sigma$ is an assignment of a three-position to each particle at each point in time.

- $m : P \to M$ is an assignment of a (determinate) mass property, such as $9.11 \times 10^{-31} kg$, to each particle. The space M of mass properties has the structure $M \sim (\mathbb{R}_0^+, +)$: that is, M is isomorphic to the nonnegative part of the real line, where 'isomorphism' is understood in the restricted sense of 'preserving addition'. (The structure of the space of mass properties is not as rich as that of the reals; in contrast to real numbers, one cannot multiply two masses to obtain a third.)

- $q_s : P \to Q_s$ is an assignment of a (determinate) charge property, such as $-1.6022 \times 10^{-19} C$, to each particle. The space Q_s of 'standard' charge properties has the structure $Q_s \sim (\mathbb{R}, +)$, i.e. Q_s is isomorphic (in the same restricted sense) to the real line. (The subscript 's' (and corresponding adjective 'standard') is for contrast with the later case of 'Feynman' charges.)

- $\mathbf{E}$ is a three-vector field — the electric field. (Formally: $\mathbf{E} : \Sigma \times T \to T\Sigma$, with $(\mathbf{E}(\mathbf{x}, t) \in T_x\Sigma)$ for all $x \in \Sigma, t \in T$.)

- $\mathbf{B}$ is another three-vector field — the magnetic field. (Formally: $\mathbf{B} : \Sigma \times T \to T\Sigma$, with $\mathbf{B}(\mathbf{x}, t) \in T_x\Sigma$ for all $x \in \Sigma, t \in T$.)

Albert's view, described in section 2.3.2, amounts to the claim that structures of this form $\mathcal{S}_{Newt}$ contain no element of conventionality; that is, that such structures 'carve electromagnetic reality at the joints'.

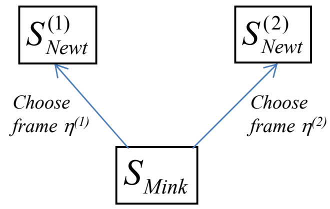
Then we noticed that we could have Lorentz invariance if we allowed $\mathbf{E}$ and $\mathbf{B}$ to transform nontrivially under Lorentz transformations; but we took it that this required regarding **E** and **B** as non-fundamental, and as defined in terms of something more fundamental only relative to a choice of frame. We therefore introduced a class of structures that (for want of a better name) we will call *Minkowski structures*, i.e. mathematical entities of the form

$$\mathcal{S}_{Mink} = \langle M, g; P, v^a, m, q, F_{ab} \rangle, \tag{2.40}$$

where:

- (M, g) is a Minkowski spacetime;
- v^a is an assignment of a four-vector (four-velocity) field to each particle (vanishing except on the particle's worldline). (Formally: $v^a : P \times M \to TM$, with $v^a(p, x) \in T_x M$ for all $p \in P, x \in M$.)
- F_{ab} is a two-form field: the Maxwell-Faraday tensor field. (Formally: $F_{ab} : M \to \Lambda T(0, 2)M$, with $F_{ab}(x) \in \Lambda T_x(0, 2)M$ for all $x \in M$.)
- Other elements of the structure are as above.

And we noted that, given a Minkowski structure, we could represent it by a Newtonian structure relative to a choice of frame η^a or, equivalently, a choice of simultaneity convention; but we recognized that the choice of frame or simultaneity convention was arbitrary, that it did not latch onto anything of metaphysical privilege, and, hence, that different Newtonian structures obtainable from the same Minkowskian structure were to be regarded as different ways of representing the same underlying reality:



But, we noticed, the idea that *Minkowski* structures were fundamental seemed to force upon us a nonstandard time reversal operation, according to which the **E** field, but not the **B** field, flips sign. *Then* we (Malament) noticed that we could recover the standard time reversal operations if we allowed F_{ab} to transform nontrivially under time reversal (specifically, if F_{ab} picked up a sign flip under time reversal); but we took it that this required regarding F_{ab} (and, in consequence, also v^a) as non-fundamental, and as defined in terms of something more fundamental only relative to a choice of temporal orientation. We therefore introduced the notion of a *Malament structure*, i.e. a mathematical entity of the form

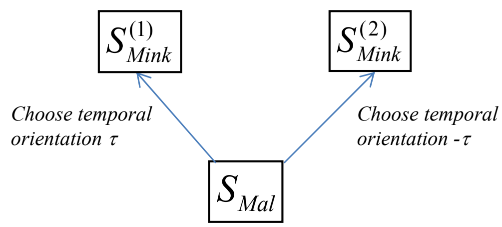
$$\mathcal{S}_{Mal} = \langle M, g; P, w_u, m, q_s, f_m \rangle, \tag{2.41}$$

where:

- P is a set of particles.
- $w_u : P \to W_u$ is an assignment of an undirected worldline to each particle. (The set W_u of undirected worldlines can be identified with the set of images of inextendible timelike curves in M.)
- $f_m : L_u \times Q_s \to TM$ is the (Malament) electromagnetic field. Here, L_u is the set of undirected tangent lines; it can be identified with the set $TM\backslash\sim$ of equivalence classes under the equivalence relation: $v^a_{(1)} \sim v^a_{(2)}$ iff $v^a_{(1)} = \lambda v^a_{(2)}$ for some $\lambda \in \mathbb{R}$. We have $f_m(l_u, q) \in T_xM$ whenever l_u is a line in T_xM.

- Other elements of the structure are as above.

We then noted that, given a Malament structure, we could represent it by a unique Minkowski structure relative to a choice of temporal orientation; but we recognized that the choice of temporal orientation was arbitrary, that it did not latch onto anything of metaphysical privilege, and, hence, that different standard Minkowskian structures obtainable from the same Malament structure were to be regarded as different ways of representing the same underlying reality:



'Feynman"s point was then that there was an alternative to Malament structures, apparently at least as defensible, although this alternative did not recover the standard time reversal operations: we could hypothesize instead that the more fundamental reality was well-represented by mathematical entities of the form

$$\mathcal{S}_{Feyn} = \langle M, g; P, w_d, m, q_f, f_f \rangle, \tag{2.42}$$

where

- $w_d : \mathbf{P} \to W_d$ is an assignment of a *directed* worldline to each possible particle. (The space W_d of directed worldlines can be identified with a set of equivalence classes of inextendible timelike curves, under the equivalence relation that relates all and only pairs of curves whose parameters increase in the same time sense as one another.)

- $q_f : \mathbf{P} \to Q_f$ is an assignment of a determinate *Feynman charge property* to each possible particle. The space Q_f of 'Feynman' charge properties has the structure $Q_f \sim (\mathbb{R}_0^+, +)$, corresponding to our earlier remark that, for Feynman, 'all charges are positive'.

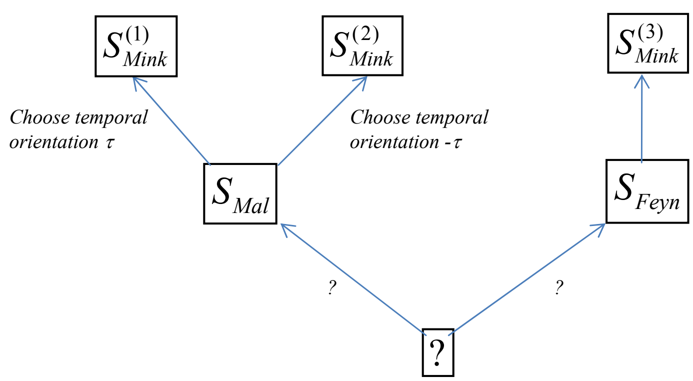
- $f_f : L_d \times Q_f \to TM$ is the (Feynman) electromagnetic field. Here, L_d is the set of directed tangent lines; it can be identified with the set $TM \backslash \sim$ of equivalence classes under the equivalence relation: $v^a_{(1)} \sim v^a_{(2)}$ iff $v^a_{(1)} = \lambda v^a_{(2)}$ for some $\lambda > 0$. We have $f_f(l_d, q) \in T_x M$ whenever l_d is a (directed) line in $T_x M$.

To complete the summary of our account thus far: We then noted that, given a Feynman structure, a representation convention can be set up according to which there is a unique standard 4D structure that represents the given Feynman structure, even *without* the selection of any conventional temporal orientation, or indeed any conventional pieces of structure:

$$\begin{tikzcd} \boxed{S_{Mink}} \\ \boxed{S_{Feyn}} \arrow[u] \end{tikzcd}$$

The 'structuralist' wants to continue this pattern: whereas the advocate of (say) the fundamentality of standard Minkowskian structures regards a large class of Newtonian structures as differing from one another only on choices of convention ('choice of frame'), not on matters of fundamental ontology (which latter are given by S_{Mink}); and whereas the advocate of the fundamentality of Malament structures regards a class of two standard Minkowskian structures as differing from one another only on choices of convention (in this case, temporal orientation), while the fundamental ontology is given by S_{Mal}; so the 'structuralist' wants to regard the elements of a class that contains *both Malament and Feynman structures* as differing from one another only on choices of convention. Malament and Feynman structures, according to the structuralist, will be equally good representors of some more fundamental underlying reality.

So far so good. It *seems*[7] reasonable, however, to require that we say more directly what the nature of this underlying reality is, rather than just 'it's something that can equally well be represented by this Malament or this Feynman structure.' That is, it seems reasonable to demand that we 'fill in the question-marks' in the following diagram:



It is in this attempt at more direct description of the nature of reality that the emphasis on relations arises. The idea is to articulate a fifth type of structure, that of 'relationist structure', to hypothesize that *that* captures the fundamental nature of electromagnetic reality better than any of the four alternatives we

[7] *Perhaps* there are limits to how far this demand can be pushed. Perhaps, that is, we eventually reach a level at which we are compelled to recognize the existence of conventionality, *but* we cannot describe the representation relations, or give a more direct description of the underlying, convention-independent reality. Interesting questions concern whether or not this happens and, if so, where it happens, and why it happens where it does.

have articulated so far, and to show how a given structure of this fifth type can be represented by a Malament, Feynman, Minkowskian or Newtonian structure relative to the selection of a certain number of arbitrary, but well-understood, conventions.

2.5.2 Relational structures

Suppose that the more fundamental story is as follows. Let M, g, P, w_u and m be (respectively) a manifold, metric, set of particles, assignment of undirected worldlines to particles, and assignment of mass properties to particles, as before. But, in place of a space of monadic charge properties (Q_m or Q_f) and an ascription (q_m, q_f respectively) of these monadic properties to particles, we have a binary *relation* $q_r : P \times P \to \mathbb{R} \cup \{\infty\}$, satisfying the following constraints:

'Reflexivity': $\forall p \in P, q_r(p, p) = 1.$

'Antisymmetry': $\forall p_1, p_2 \in P, q_r(p_1, p_2) = q_r(p_2, p_1)^{-1}.$

'Transitivity': $\forall p_1, p_2, p_3 \in P, q_r(p_1, p_2) \cdot q_r(p_2, p_3) = q_r(p_1, p_3);$

Heuristically: in terms of Malament structures, q_r corresponds to a 'charge ratio' relation; while, in terms of Feynman structures, the absolute value of q_r corresponds to the charge ratio, while the sign of q_r encodes whether or not the worldlines of the two particles have the same temporal direction as one another. But it is crucial to note that neither of these translation schemata forms part of the relationist account *per se*. According to the relationist, there is just q_r.

We are then dealing with *relational structures*: mathematical entities of the form

$$S_{rel} = \langle M, g; P, w_u, m, q_r, f_r \rangle, \tag{2.43}$$

where

- q_r is as above.

- $f_r : L_u \times P \to TM$ is a map assigning a four-vector in T_xM to every pair (l_u, p) such that l_u is a line in T_xM (for some $x \in M$).

- Other elements of the structure are as above.

A relational structure represents an electromagnetic world as containing point particles $p \in P$ that have monadic mass properties[8], and that bear a 'charge-ratio'-like *relation* to one another; the electromagnetic field is accordingly reconceived as f_r rather than f_m or f_f, so that it makes no reference to monadic charge properties.

2.5.3 Malament and Feynman structures as conventional representors of a relational reality

We now wish to explore the ('structuralist') suggestion that it is the relational structures that best 'carve electromagnetic reality at its joints', and that Malament and Feynman structures arise as convenient mathematical tools which, however, require us to make some choices of arbitrary convention that need not be made by the pure relational approach. Specifically, we wish to explore the nature of the representation relation between (represented) relational structures and (representing) Malament or Feynman structures.

The following definition will prove useful: Say that a particle $p \in P$ *has zero charge* iff for some $p' \in P$, $q_r(p, p') = 0$.[9]

Suppose, then, that we are given a relational structure, i.e. an entity of the form (2.43). We first wish to represent this via a Malament structure. To do so,

---

[8]A more thorough-going structuralism, of course, would treat mass, as well as charge, in a relational way. We omit this complication for brevity.

[9]This definition has the consequence that if, intuitively, *all* particles have zero charge, none will count as having zero charge according to the definition. This consequence is unwanted, but does not create any problems. In such cases, the indifference of the particles to the EM field will be encoded in f_r (which would everywhere take zero vectors as its values).

we proceed as follows:

1. Let Q_m be a space with the structure $Q_m \sim (\mathbb{R}, +)$. (This structure suffices to define a notion of multiplication by an arbitrary real number on Q_m.)

2. Define a function $q_m : P \to Q_m$ as follows:

   - Choose arbitrary $\tilde{p} \in P$ such that $\tilde{p}$ has nonzero charge. (The existence of some such particle, providing that P is nonempty, is guaranteed by the axioms governing q_r; cf. footnote 9. If P is empty, then, of course, any function with domain P is trivial.)
   - Choose arbitrary nonzero charge $\tilde{q} \in Q_m - \{0\}$.
   - Define a function $q_m : P \to Q_m$ as follows:

     (a) $q_m(\tilde{p}) = \tilde{q}$.

     (b) For all $p' \in \mathbf{P_r}, q_m(p') = q_m(p) \cdot q_r(p', p)$.

3. Define a map $f_m : L_u \times Q_m \to TM$ as follows:

$$\forall l_u \in L_u, \forall q \in Q_m, f_m(l_u, q_m) = \frac{q}{\tilde{q}} f_r(\langle l_u, \tilde{p} \rangle). \tag{2.44}$$

4. Form the Malament structure $\langle M, g; P, w_u, m, q_m, f_m \rangle$.

We note that, given a relational structure, we have the following arbitrary choice of convention to make, in order to determine the Malament structure that would represent it: the charge $q_m(\tilde{p}) \in Q_m - \{0\}$ for an arbitrarily selected charged particle $\tilde{p}$.

To represent our given relational structure using a Feynman structure, on the other hand, we would proceed as follows:

1. Let Q_f be a space with structure $Q_f \sim (\mathbb{R}_0^+, +)$. (This structure suffices to define a notion of multiplication by an arbitrary nonnegative real number on Q_f.)

2. If all particles in $\mathbf{P_r}$ have zero charge, set $q_m(p) = 0 \in Q_f$, for all $p \in \mathbf{P_r}$. If some particle in $\mathbf{P_r}$ has nonzero charge, then:

   - Choose arbitrary $\tilde{p} \in P$ with nonzero charge.
   - Choose arbitrary nonzero charge $\tilde{q} \in Q_f - \{0\}$.
   - Define $q_f : P \to Q_f$ as follows:

     (a) $q_f(\tilde{p}) = \tilde{q}$.

     (b) For all $p' \in P, q_f(p') = q_f(p) \cdot |q_r(p', \tilde{p})|$.

3. Construct the ascription w_d of *directed* worldlines to particles, as follows. First, note that the set W_d has two natural pieces of structure. (i) If $w_1, w_2 \in W_d$, say that w_1 *is codirected with* w_2 iff w_1 'points in the same temporal direction as' w_2.[10] Codirectedness is then an equivalence relation on W_d, partitioning W_d into two mutually exclusive and jointly exhaustive classes. (ii) If $w_3, w_4 \in W_d$, *or* if $w_3 \in W_d$ and $w_4 \in W_u$, say that w_3 *is coextensive with* w_4 iff w_3, w_4 occupy the same set of points of M. Coextensiveness (in the first sense) is also an equivalence relation on W_d, this time partitioning W_d into uncountably many equivalence classes of two elements each. Then:

   - Select an arbitrary directed worldline w that is coextensive with the undirected worldline $w_u(p)$ that our relational structure ascribes to p; let $w_d(p) = w$.

[10]Rigorously: $w_1 \sim w_2$ iff, for any continuous nowhere-vanishing timelike vector field τ^a on M and any $s_1, s_2 \in \mathbb{R}$,

$$\left(\eta_{ab}\left(\frac{dw_1}{ds}\right)^a |_{s_1}\tau^b(w_1(s_1))\right)\left(\eta_{cd}\left(\frac{dw_1}{ds}\right)^c |_{s_2}\tau^d(w_1(s_2))\right) > 0. \tag{2.45}$$

- For all other particles $p' \in \mathbf{P}$:
  - If $q_r(p', p) > 0$, let $w_d(p')$ be the unique element of W_d that is coextensive with $w_u(p')$ and codirected with $w_d(p)$.
  - If $q_r(p', p) < 0$, let $w_d(p')$ be the unique element of W_d that is coextensive with $w_u(p')$ and *not* codirected with $w_d(p)$.

4. Define a map $f_f : L_d \times Q_f \to TM$ as follows:

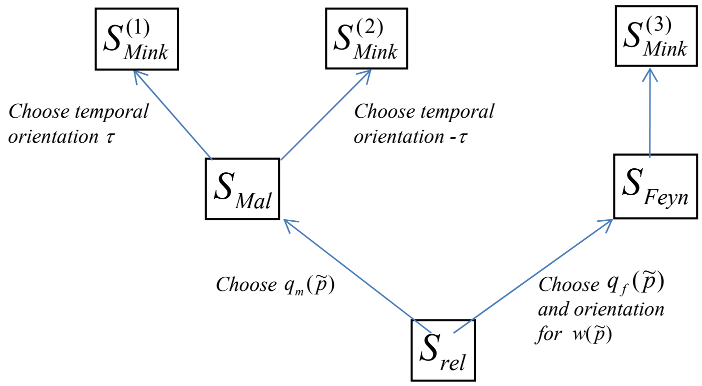
$$\forall l_d \in L_d, \forall q \in Q_f, f_f(l_d, q) = \pm \frac{q}{\tilde{q}} f_r(l_d, \tilde{p}), \tag{2.46}$$

where the positive sign applies iff the orientation of l_d is the same as that of $w_d(\tilde{p})$.

5. Form the Feynman structure $\langle M, g; P, w_d, m, q_f, f_f \rangle$.

In this case, we had to make *two* arbitrary choices of convention: the charge $q_f(\tilde{p}) \in Q_f - \{0\}$ of our arbitrarily selected charged particle $\tilde{p}$, and the orientation of its worldline. The superficial appearance that this involves 'more conventionality' than does the construction of a Malament from a relational structure, however, is no more than that: on any reasonable way of quantifying 'degree of conventionality', the selection of an arbitrary element of $Q_m \sim \mathbb{R}$ will count as the introduction of 'just as much convention' as will the selection of an arbitrary element of $Q_f \sim \mathbb{R}_0^+$ and an arbitrary orientation for a given worldline.

To sum up our structuralist program, then: we have written down prescriptions for constructing Malament and Feynman structures from a given relational structure $\langle M, g; P, w_u, m, q_r, f_r \rangle$. In this way, it can be a consequence of our third candidate ontology, according to which it is the relational structures that best 'carve electromagnetic reality at the joints', that the choice between representation via a Malament structure and representation via a Feynman structure is merely a choice of convention:



Thus (given our earlier accounts of how the ontologies on which Malament and Feynman structures are based give rise to distinct geometrical time reversal operations), we have shown how a relationist can support the claim that answers to questions like whether or not four-velocity flips sign under time reversal, whether time reversal turns particles into antiparticles, and so on, are convention-dependent: questions that have no determinate answers until we implicitly choose our convention (by answering the question, or otherwise).

2.6 Conclusions and open questions

In this final section, we summarize our conclusions to date, and then indicate some open issues that we would like to resolve.

Summary of conclusions from this paper.

We have articulated the 'geometric' notion of time reversal implicit in Malament's work, according to which time reversal consists in leaving all [other] fundamental quantities alone, and merely flipping the temporal orientation. This allows us to give an account, as the passive and active notions of time reversal cannot, of how it may come about that a coordinate-independent quantity such as F^{ab} transforms nontrivially under time reversal; and it is in any case the right notion to focus on if one's interest is in the conclusions about the structure of spacetime that can legitimately be drawn from the invariances of our best theories. We have then discussed four approaches to time reversal in classical electromagnetism in the light of this geometric conception: Albert's, Malament's, the 'Feynman' approach, and the structuralist approach. Only according to Albert is the theory not time reversal invariant; we have rejected Albert's account by appeal to Ockham's Razor.

Theory choice.

This does, however, leave us with an apparent case of underdetermination: how might one choose between the Malament, Feynman and Structuralist ontologies, and which seems to be preferable all things considered? We are not sure how best to answer this question; so let us merely list several considerations that *may* tell one way or another.

Firstly: one feeling is that Structuralism is preferable because it eliminates distinctions that seem to be devoid of differences. But it would be better if this 'feeling' could be replaced with argument, and it is difficult to turn the sentiment expressed in the preceding sentence into an argument for structuralism without falling foul of the point that the choice between structuralism and its alternatives is itself a choice that is, in a very similar sense, 'underdetermined by the physics'.

Secondly: it is not clear that the Feynman account can give a reasonable treatment of *neutral* particles. We skirted over this difficulty in our above discussion, but it is not hard to see, particularly in the context of the attempt to define a Feynman structure to represent a relational reality: in the case of a neutral particle, there is nothing 'in the physics' to determine what the orientation of the particle's worldline should be. To insist that even in this case there must nevertheless be a fact about the worldline's orientation seems ontologically extravagant;

to treat neutral particles in Malament's way, while retaining a Feynman treatment of charged particles, though, seems to amount to adopting an ugly hybrid position.

Thirdly: it is not clear that the Structuralist account can give a reasonable treatment of the electromagnetic field, either in cases in which all particles are neutral, or in cases in which there are no particles at all (i.e. cases of *vacuum* solutions of the Maxwell equations). Let us take up the second point first. The point here is that if the relationist electromagnetic field f_r *just is* a map from $L_u \times P$ to TM then, if P is empty, f_r is a map with empty domain; thus, the structuralist account does not seem to have the resources to underwrite a genuine physical difference between any one vacuum solution and any other. Going back to the case of neutral particles: similarly, in any case in which all actual particles are neutral, the relationist electromagnetic field must assign the zero four-vector to every pair (l_u, p); thus, again, it cannot underwrite genuine physical differences between solutions of the Maxwell equations that differ radically on the value of the Maxwell-Faraday field F_{ab}. Of course, the structuralist could bite the bullet and say that, indeed, there is no genuine physical difference between such pairs of solutions; whether or not this (bullet-biting) move would lead to trouble is an open question.

Fourthly: the Malament account does not seem to sit particularly well with the idea that, at a rather fundamental level, the Maxwell-Faraday tensor is to be thought of as the curvature of a $U(1)$ connection one-form A_a. If one takes this latter idea seriously, one *seems* to be led to something like the Feynman view: the most fundamental representation of the electromagnetic field is (according to this idea) as a two-form, *not* as a map from tangent lines and either charges or particles to four-vectors. Thus, connection realism seems to lead to the Feynman view of time reversal by default.

One final (and very plausible) possibility is that the underdetermination in question simply cannot be correctly resolved within the confines of classical electromagnetism, and that it is only by viewing classical electromagnetism as the classical limit of a quantum field theory, and thus obtaining further ontological insight as to the nature of charged particles and of the electromagnetic field, that one runs across considerations that favor the true ontological position over others. The investigation of these possibilities is a future project.

Conventionality of spacetime structure?

An intriguing issue arises on the supposition that structuralism is indeed correct. In that case, as we have emphasized, the difference between the Malament and Feynman languages is just that — a difference in language; one's choice of language is a convention. In the case of classical electromagnetism, nothing of ontological substance even *threatens* to hang on the choice of convention; in particular, the existence or nonexistence of a preferred temporal orientation does not, since the theory comes out time reversal invariant according to both Malament and Feynman. A more interesting case would be one, if any such there be, in which the time reversal invariance of the theory was (according to structuralism) a convention-dependent matter. Given the standard link between spacetime symmetries and spacetime structure, this would render the question of whether or not a privileged temporal orientation exists a convention-dependent matter. It is not immediately clear whether or not this makes sense. If it does, the details have yet to be worked out; if it does not, this seems to be a strong argument against the structuralist position.

Field theory.

The original motivation for this project was the feeling that the existence of a CPT theorem is rather puzzling — why should *charge conjugation* be so intimately related to spacetime symmetries? The point here is that, according to the 'Feynman' proposal, the operation that *ought* to be called 'time reversal' — in the sense that it bears the right relation to spatiotemporal structure to deserve that name — is the operation that is usually called TC; on this proposal, the theorem known as the 'CPT theorem' would be more properly called a PT theorem, and (the thought continues) perhaps this opens the door to new insights into why that theorem should hold. There are arguments for the usual identifications of certain quantum-field-theoretic operators as time reversal, charge conjugation and parity reversal; but those arguments invoke tacit assumptions such as that four-velocities are always future-directed, and we have seen that in the classical case (at any rate), there is nothing incoherent about denying such assumptions. A future project is to investigate a geometrical understanding of the (classical and quantum field-theoretic) 'CPT' theorems, drawing on this suggestion.

Chapter 3

A classical PT theorem

3.1 Introduction

The CPT theorem of quantum field theory states that any quantum field theory invariant under the restricted Lorentz group is also invariant under CPT, the composition of charge conjugation, parity reversal and time reversal.

This theorem is usually thought of as a peculiarly *quantum*-theoretic result. But, on reflection, this is *prima facie* odd. One normally expects a quantum theory and its classical counterpart to have the same symmetries as one another. The known modes of exception to this normal expectation are (i) cases of spontaneous symmetry breaking, in which the dynamics but not the vacuum state of the quantum field theory have all the symmetries of the classical theory, and (ii) cases of so-called anomalous symmetry breaking, in which the quantum theory fails to have some classical symmetry because the regularization breaks the symmetry in question. (See, for example, (Peskin & Schroeder, 1995, chapters 11 and 19), for an exposition of the phenomena of spontaneous and anomalous symmetry breaking (respectively).) We notice that both of these provide ways for the quantum field theory to have *fewer* symmetries than its classical counterpart. *If* the CPT theorem is indeed peculiarly quantum-theoretic, it seems to be one of a kind: our sole example of a symmetry that classically can be absent, but quantum-theoretically is (somehow) guaranteed.

Our sense of puzzlement is deepened when we look at proofs of the CPT theorem. The mathematical heart of extant proofs seem to be such facts as that the 'PT' transformation is connected to the identity in the complexification of the Lorentz group; thus one can prove CPT theorems via analytic continuation arguments. But there is no apparent reason why analytic continuation arguments should be any less applicable to classical than to quantum field theories. One starts to suspect that, in fact, there is *nothing* essentially quantum-theoretic about the CPT theorem, and that it is no more than historical accident that it was discovered in the context of quantum as opposed to classical field theory.

We issue to ourselves, therefore, the following disjunctive challenge: *either* identify the key difference between the structure of classical and quantum theories in virtue of which a quantum CPT theorem exists while no classical counterpart does, *or* provide an analogous classical theorem.

Standard quantum field theories deal with two types of dynamical fields: tensor fields, and spinor fields. Surprisingly, it turns out that we must take up different disjuncts of our challenge in the tensor and spinor cases.

For theories whose dynamical fields are tensor fields, it turns out to be possible to prove a classical analogue of the CPT theorem. This is the main task of the present paper. The claim we will prove is that for a particular class of such theories, any theory that is restricted-Lorentz-invariant is also PT-invariant. The theorem we shall prove is not entirely new. A very similar result was stated, and a proof sketched, already in a 1955 paper by John S. Bell (Bell, 1955). The present paper draws heavily on Bell's work, and is in large part an exegesis of that work. However, the statement of the theorem, and the proof-sketch, that Bell provides in that work is rather brief. Here we state the theorem explicitly and set out a proof (differing slightly from Bell's own) in full detail.

For *spinor* field theories, things appear to be otherwise. Below, we will give two strong reasons for thinking that there is no classical analogue of the CPT theorem in the spinor case. If we accept that indeed there is not, this raises the urgent question of how, in *this* case, the move to the quantum theory brings with it a guarantee of CPT invariance. This is, however, a question to which we have as yet no answer to offer. Thus, the brief discussion of spinors in this paper functions only to suggest a research project.

The structure of the paper is as follows. In section 3.2, we review the definitions of some of the concepts that our discussion will require, and we establish two lemmas that will be required for the proof of our theorem. Section 3.3 states and proves the theorem itself. Section 3.4 is a brief discussion of the implications and limitations of the theorem. Section 3.5 states the reasons for thinking that there is no analogue of this theorem for spinor field theories. Section 3.6 is the conclusion.

3.2 Definitions and lemmas to be invoked

3.2.1 Definitions

We presuppose familiarity with such standard concepts of differential geometry as that of a (differentiable) manifold, diffeomorphism, tangent and cotangent space, push-forward and pull-back of a diffeomorphism, and fibre bundle. (For details of these see, for example, (Isham, 1999).) The following definitions are stated in order to fix notation and terminology.

Tensor fields.

Let M be a manifold. Let T_pM, T_p^*M be (respectively) the tangent and cotangent spaces at a point $p \in M$. Then, a *tensor of type (r,s)* at a point $x \in M$ is an element of the tensor product space

$$T_x^{p,q}M := (\otimes^p T_x M) \otimes (\otimes^q T_x^* M) \,. \tag{3.1}$$

Let $T^{(p,q)}M$ be the fibre bundle over M with the property that, for all $x \in M$, the fibre over x is the space $T_x^{(p,q)}M$.

A *tensor field* of type (p, q) on M is a cross-section of the bundle $T^{(p,q)}M$.

Lorentz group.

Let M be a four-dimensional differentiable manifold that is diffeomorphic to $\mathbb{R}^4$. Let η be a flat metric on M with 'Lorentzian' signature $(1,3)$. Let *Diff*(M) be the group of diffeomorphisms of M. Let the *diffeomorphism Lorentz group* $L \subset$ *Diff*(M) be the subgroup consisting of those diffeomorphisms that leave η invariant: that is,

$$\forall h \in \mathit{Diff}(M), l \in L \text{ iff } l * \eta = \eta. \tag{3.2}$$

Let $\tilde{\eta}$ be the diagonal 4×4 matrix $diag(-1,1,1,1)$. Let the *matrix Lorentz group* be the group $\tilde{L} \subset M(4,\mathbb{R})$ of 4×4 real matrices possessing the property that, for all $m \in M(4,\mathbb{R})$, $m \in \tilde{L}$ iff $m \cdot \tilde{\eta} \cdot m^T = \tilde{\eta}$, where $\cdot$ denotes matrix multiplication.

Tensor representations of the Lorentz group.

Given any tensor bundle $T^{(p,q)}M$ over a manifold M that is equipped with a flat Lorentzian metric η, we can define the *tensor representation* $R^{(p,q)}$ *of type* (p,q) of the Lorentz group $L \subset$ *Diff*(M) on $T^{(p,q)}M$, as follows.

Let $v^i : i = 1,\ldots,4$, $w_j : j = 1,\ldots,4$ be bases for T_xM, T_x^*M respectively. Suppose $t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q} \in T_x^{(p,q)}M$. (That is, $t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}$ is a tensor of type (p,q) at the point $x \in M$. Note that here we are using the 'abstract index notation': $t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}$ is a *tensor*, not a component of a tensor relative to some coordinate system, and the $\{a_i\}, \{b_j\}$ are abstract indices rather than variables ranging over integers.) Then, $t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}$ can be expanded as follows:

$$t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q} = \sum_{i_1,\ldots,i_p,j_1,\ldots,j_q=1,\ldots,n} a_{i_1 \ldots i_p j_1 \ldots j_q} \left(v^{i_1} \otimes \ldots \otimes v^{i_p} \otimes w_{j_1} \otimes \ldots \otimes w_{j_q} \right),$$

where the $a_{i_1 \ldots i_p j_1 \ldots j_q} \in \mathbb{R}$ are coefficients. For any vector $v \in T_xM$, covector

$w \in T_x^*M$ and Lorentz transformation $l \in L$, let l^*v, l_*w denote (respectively) the push-forward of v and the pull-back of w by the diffeomorphism l. Then, define the representation $R^{(p,q)}$ of L on $T^{(p,q)}M$ as follows: for all $l \in L$, all $x \in M$ and all $t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q} \in T_x M$,

$$\begin{aligned}
&(R^{(p,q)}(l))(t^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}) \\
=&\sum a_{i_1 \ldots i_p j_1 \ldots j_q} \left(l^* v^{i_1} \otimes \ldots \otimes l^* v^{i_p} \otimes l_* w_{j_1} \otimes \ldots \otimes l_* w_{j_q}\right) \\
\in&T^{(p,q)}_{l(x)} M.
\end{aligned}$$

With slight abuse of notation, we also write $R^{(p,q)}$ for the induced representation of l on the space of *cross-sections* of $T^{(p,q)}M$, as follows. Let ψ be such a cross-section. Then, we write

$$\left(\left(R^{(p,q)}(l)\right)(\psi)\right)(x) = \left(R^{(p,q)}(l)\right)\left(\psi\left(l^{-1}(x)\right)\right). \tag{3.3}$$

3.2.2 Lemmas to be invoked

The proof to be offered in section 3.3 will invokes Lemma 2, below. Lemma 1 is required for the proof of Lemma 2.

Our first lemma is an elementary theorem of complex analysis:

Lemma 1. *Let $f : \mathbb{C} \to \mathbb{C}$ be an everywhere complex-differentiable function of a single complex variable. Suppose that f vanishes on the real line. Then, f vanishes on all of $\mathbb{C}$.*

Proof. Since f is everywhere complex-differentiable, it has a Taylor expansion

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} \theta^n$$

that converges to f for all $\theta \in \mathbb{C}$. If $f(\theta) = 0$ for all real θ, we must have

$f^{(n)}(0) = 0$ for all n. But then $f(\theta) = 0$ for all $\theta \in \mathbb{C}$. □

We use this to establish a further result that we will require in section 3.3:

Lemma 2. *As above, let $\tilde{L}$ be the matrix Lorentz group. Let $\tilde{J} : \tilde{L} \to \mathbb{R}$ be a function on $\tilde{L}$ that is polynomial in the matrix entries. Suppose that $\tilde{J}$ vanishes on the connected component $\tilde{L}_+^\uparrow$ of $\tilde{L}$. Then, we also have $\tilde{J}(diag(-1,-1,-1,-1)) = 0$.*

Proof. Let $D \subset \mathbb{C}$ be given by $D := \mathbb{R} \cup \mathbb{R} + i\pi$. Let $p : D \to \tilde{L}$ be given by

$$p(\theta) = \begin{pmatrix} \cosh\theta & -\sinh\theta & 0 & 0 \\ -\sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.4}$$

Then, any function $\tilde{J}$ on $\tilde{L}$ induces a function j on D, via composition:

$$j(\theta) := \tilde{J}(p(\theta)). \tag{3.5}$$

Since $\tilde{J}$ vanishes on $\tilde{L}_+^\uparrow$, and $p(\theta) \in \tilde{L}_+^\uparrow$ for all real θ, we have that j vanishes on $\mathbb{R}$. But also, if $\tilde{J}$ is a polynomial in the matrix entries, then j is a polynomial in $\{\cosh\theta, \sinh\theta\}$, in the sense that there exists a non-negative integer p and real coefficients $a_{mn} : m, n = 0, \ldots, p$ such that for all $\theta \in D$,

$$j(\theta) = \sum_{m,n=0}^{p} a_{mn} \cosh^m\theta \sinh^n\theta. \tag{3.6}$$

But any function on a subset $D \subseteq \mathbb{C}$ that is polynomial in $\{\cosh\theta, \sinh\theta\}$ in this sense clearly has an everywhere complex-differentiable extension to all of $\mathbb{C}$ (namely, the function given by allowing the variable θ in the polynomial expression (3.6) to range over all of $\mathbb{C}$). Hence, it follows from Lemma 1 that since j vanishes on $\mathbb{R}$, j vanishes on $\mathbb{R} + i\pi$ also. In particular, $j(i\pi) = 0$; so, $\tilde{J}(p(i\pi)) = 0$. But, $p(i\pi) = diag(-1, -1, -1, -1)$. Hence, we have shown that $\tilde{J}(diag(-1, -1, -1, -1)) = 0$, as claimed.

$\square$

3.3 Classical PT theorem for tensor field theories

We are now ready to state and prove our main theorem.

In outline, the result is as follows. (The contents of this paragraph will all be stated more rigorously below.) We consider a classical theory given by a system of partial differential equations (PDEs) on a specified set of fields. Let Φ be the space of kinematically allowed fields. (In the general case, we may be dealing with a theory containing a number of interacting fields — scalar fields, tensor fields, etc — so, for a given theory, an element of Φ will be an ordered m-tuple of specified numbers of scalar fields, vector fields, rank 2 tensor fields, etc.) We note that any PDE can be expressed as the vanishing of some functional $F : \Phi \to \mathbb{R}^M$ of the fields. (That is, F encodes the dynamics in the sense that: $\phi \in \Phi$ is dynamically allowed iff $F(\phi)$ is the zero map on M.) We *assume that F is a 'local' polynomial in the fields and their derivatives.* It can then be proved that, if the set S of solutions of the equation $F(\phi) = 0$ is invariant under the tensor representation of the restricted Lorentz group $L_+^\uparrow$, then S is actually invariant under the tensor representation of the whole of the proper Lorentz group L_+ (i.e. including total reflections as well as rotations and boosts).

More rigorously, we have the following

Theorem 1 (Classical PT theorem for polynomial systems of real tensor fields.)**.** *Let M be a differentiable manifold that is diffeomorphically $\mathbb{R}^4$.*

Let Φ be a space of n-tuples of tensor fields of specified types on M. That is: For some fixed integer $n > 0$ and functions $p : \{1, \ldots, n\} \to \mathbb{N}, q : \{1, \ldots, n\} \to \mathbb{N}$, *let Φ be the set of n-tuples $\phi \equiv (\phi_1, \ldots, \phi_n)$, where, for each $i \in \{1, \ldots, n\}$, ϕ_i is a tensor field of type $(p(i), q(i))$ over M.*

Let η be a flat Lorentzian metric on M. Let L be the group of manifold diffeomorphisms $l : M \to M$ leaving η invariant (i.e., L is the Lorentz group). Let $L_+^\uparrow, L_-^\downarrow, L_-^\uparrow, L_+^\downarrow$ be the connected subsets of L that reverse neither time sense nor parity, time sense but not parity, parity but not time sense, and both time sense and parity respectively.

For all $p, q \in \{0, 1, \ldots\}$, let $R^{(p,q)}$ be the tensor representation of type (p, q) of the Lorentz group L on the space of tensor fields of type (p, q). Let R^T be a representation of L on Φ whose restriction to the proper[1] Lorentz group L_+ coincides with that naturally induced by the $\{R^{(p,q)}\}$: that is, let R^T be such that for all $\phi \in \Phi$, $l \in L_+$, we have

$$\left(R^T(l)\right)(\phi) = \left(\left(R^{(p(1),q(1))}(l)\right)(\phi_1), \ldots, \left(R^{(p(n),q(n))}(l)\right)(\phi_n)\right). \tag{3.7}$$

Let $F : \Phi \to \mathbb{R}^M$ be a functional that is 'polynomial in the fields ϕ_i and their derivatives', in the following sense: there exists an inertial coordinate system $x : M \to \mathbb{R}^4$, non-negative integers p, q and real coefficients $\{a_{m_1,\ldots,m_q} \in \mathbb{R} : m_1, \ldots, m_q = 0, \ldots, p\}$ such that for all $\phi \in \Phi$,

$$F(\phi) = \sum_{m_1,\ldots,m_q=0}^{p} a_{m_1,\ldots,m_q} (\psi_1)^{m_1} (\psi_2)^{m_2} \ldots (\psi_q)^{m_q}, \tag{3.8}$$

where each ψ_j is a specified partial coordinate derivative (possibly zeroth order) of a specified one of the ϕ_i relative to the chart μ, and multiplication is defined

[1] That is: we are stipulating that 'our fields transform as true tensors under proper Lorentz transformations', but we are saying nothing about 'how they transform under improper Lorentz transformations', i.e. transformations that reverse exactly one of time sense and parity. In particular, we are not ruling out so-called 'pseudo-tensor' fields that 'pick up an additional sign flip relative to true tensors' under improper Lorentz transformations. The distinction between true tensors and pseudo-tensors is irrelevant for present purposes, since the theorem to be proved deals only with proper Lorentz transformations.

pointwise in the obvious way.[2]

Let $S \subseteq \Phi$ be given by

$$S := \{\phi \in \Phi : F(\phi) = 0\} \tag{3.9}$$

[the intended interpretation being that S is the set of solutions to the partial differential equation expressed by the condition $F = 0$].

Suppose that S is invariant under $R^T(L_+^{\uparrow})$, i.e. for any $\phi \in S$ and $l \in L_+^{\uparrow}$, $(R^T(l))(\phi) \in S$ also. Then, S is actually invariant under all of $R^T(L_+)$.

Proof. The structure of the proof is as follows. In Step 1 we construct, from the functional F, a family of functions $H_\phi : \tilde{L}_+ \to \mathbb{R}^M$ on the diffeomorphism proper Lorentz group. We note that since S is $L_+^{\uparrow}$-invariant, these functions H_ϕ vanish on $L_+^{\uparrow}$. In Step 2 we derive a constraint on the form of the H_ϕ, from our assumption that F is polynomial in the fields and their derivatives. In Step 3, we combine the results of Step 2 and Lemma 2 (above) to infer that, for all $\phi \in S$, H_ϕ also vanishes at at least one point $l \in L_+^{\downarrow}$. It will follow immediately from this that S is invariant under the representatives of the whole of the proper Lorentz group, $R^T(L_+)$.

Step 1. For arbitrary $\phi \in \Phi$, define $H_\phi : L_+ \to \mathbb{R}^M$ as follows: for all $l \in L_+$,

$$H_\phi(l) = F\left[\left((R^T)(l)\right)(\phi)\right]. \tag{3.10}$$

We note, for later use, that we will have $H_\phi(l) = 0$ for all $\phi \in S$ iff S is invariant under $R^T(l)$. (This is the case because, by the definition of $S \subset \Phi$, the condition $(\forall \phi \in S)(F\left[\left(R^T(l)\right)(\phi)\right] = 0)$ is equivalent to the condition $(\forall \phi \in$

[2] We note that if this condition holds relative to one global coordinate chart on M, it holds relative to all of them. Hence, the 'polynomial' character of F does not pick out any privileged proper subset of coordinate systems.

$S)((R^T(l))(\phi) \in S).)$

Step 2. Let $\mu : M \to \mathbb{R}^4$ be a global inertial coordinate chart on M. Relative to μ, there is a privileged isomorphism between the diffeomorphism Lorentz group $L \subset \mathit{Diff}(M)$ and the matrix Lorentz group $\tilde{L} \subset M(4, \mathbb{R})$ of 4×4 matrices m having the property that $m \cdot \tilde{\eta} \cdot m^T = \tilde{\eta}$, where $\tilde{\eta} = \mathit{diag}(-1, 1, 1, 1)$ and $\cdot$ denotes matrix multiplication. Write $i : L \to \tilde{L}$ to denote this isomorphism. Hence, relative to μ, any function j L induces a function $\tilde{j}$ on $\tilde{L}$ with the same target space, in an obvious way:

$$\forall \tilde{l} \in \tilde{L}, \tilde{j}(\tilde{l}) = j(i^{-1}(\tilde{l})). \tag{3.11}$$

With slight abuse of notation, we also write $H_\phi : \tilde{L} \to \mathbb{R}^M$ for the function induced on $\tilde{L}$ by $H_\phi : L \to \mathbb{R}^M$.

We claim that, for arbitrary $\phi \in \Phi$, $\tilde{H}_\phi : \tilde{L} \to \mathbb{R}^M$ is a polynomial in the matrix entries. That is, we claim that for each $\phi \in \Phi$, there exists an integer $s > 0$ and a family of coefficient functions $a^\phi_{m_{00},\ldots,m_{33}} : M \to \mathbb{R}$ such that, for all $\tilde{l} \in \tilde{L}$,

$$\tilde{H}_\phi(\tilde{l}) = \sum_{m_{00}, m_{01}, \ldots, m_{33} = 0, \ldots, s} a^\phi_{m_{00},\ldots,m_{33}} \prod_{\mu,\nu=0}^{3} \left(\tilde{l}^\mu{}_\nu\right)^{m_{\mu\nu}}. \tag{3.12}$$

This follows from the fact that F is a polynomial in the fields and their derivatives in the sense of equation (3.8), together with the fact that R^T is the tensor representation.

Step 3. We now show that if $\tilde{H}_\phi$ vanishes on $\tilde{L}^\uparrow_+$, it also vanishes at the point $\mathit{diag}(-1, -1, -1, -1) \in \tilde{L}^\downarrow_+$.

Choose arbitrary $x \in M$. Let $\tilde{H}_{\phi,x} : \tilde{L}_+ \to \mathbb{R}$ be defined by

$$\forall \tilde{l} \in \tilde{L}_+, \tilde{H}_{\phi,x}(l) = (\tilde{H}_\phi(\tilde{l}))(x). \tag{3.13}$$

Then, $\tilde{H}_{\phi,x}$ is a polynomial in the matrix entries in the sense of Step 2, and vanishes on $\tilde{L}_+^\uparrow$ by the assumption that S is $L_+^\uparrow$-invariant. It therefore follows from Lemma 2 that $\tilde{H}_{\phi,x}(diag(-1,-1,-1,-1)) = 0$ also. But this holds for arbitrary $x \in M$; hence, we have shown that $\tilde{H}_\phi(diag(-1,-1,-1,-1))$ is the zero map on M. Since this holds for arbitrary $\phi \in S$, we have shown that if $F(\phi) = 0$ then $F\left[R^T(i^{-1}(diag(-1,-1,-1,-1)))(\phi)\right] = 0$ also; that is, we have shown that S is PT-invariant. But if a set $Q \subset \Phi$ is invariant under $R^T(L_+^\uparrow)$ and also invariant under $R^T(PT)$, then, by the facts that R^T is a representation and that any total-reflection Lorentz transformation $l \in L_+^\downarrow$ can be expressed as the product of some restricted Lorentz transformation and PT, Q is invariant under all of $R^T(L_+^\uparrow)$. Hence, S is invariant under $R^T(L_+)$, as was to be shown.

□

3.4 Limitations of the classical PT theorem and projects suggested

Restrictions required on the dynamical equations.

The theorem stated in section 3.3 treats only the case in which the dynamics is given by a set of *polynomials* in the fields and their derivatives. This assumption of polynomiality might be regarded as undesirably restrictive from the point of view of physics. For example, $\sin\phi$ and $\sqrt{\phi}$ are not polynomial in ϕ, but the essential use of trigonometric functions and square root operations in dynamical equations does not seem to be physically implausible.[3]

In fact, the proof of section 3.3 will go through under weaker restrictions on F than that it be polynomial. For example, it suffices to assume that F is built from

[3] I am grateful to Robert Geroch and Michael Kiessling (respectively) for suggesting these examples.

the dynamical fields and their derivatives by operations that are representable by power series with infinite radius of convergence. Thus, we can permit the appearance of operators such as sin and cos in the dynamical equations, and still prove the PT result.[4] However, arguably, this is still not as weak an assumption as would entitle us to the claim that we have proved a PT theorem for 'all physically reasonable' classical field theories.

It is worth comparing the theorem proved in section 3.3 with standard proofs of the quantum-theoretic CPT theorem, on this point. In the quantum case, one finds apparently wildly different proofs in the Lagrangian, axiomatic and algebraic approaches to quantum field theory. (See, for example, (Itzykson & Zuber, 2005, pp.157-9) for a proof-sketch within the Lagrangian framework, (Streater & Wightman, 1964, section 4.3) for a proof within the axiomatic approach, and (Borchers, 2000, section IV) for a 'purely algebraic' proof.) It is the Lagrangian approach that provides quantum field theories that one can regard as quantizations of particular classical field theories; in the proof of the CPT theorem that can be given within this approach, the *assumption* that the Lagrangian is polynomial must be made, just as in the classical case. In (for example) the axiomatic approach, on the other hand, it is *proved* that the so-called 'Wightman functions' are complex-analytic in the sense required for the CPT result to go through, on the basis of abstract assumptions such as the spectral condition. In this case there seems to be an important disanalogy between the classical and quantum theorems: to get the classical theorem we impose, with possibly dubious physical motivation, a condition that seems to be forced in the quantum case.

It is then interesting to ask what happens to a classical field theory that is sufficiently far from being polynomial to be restricted-Lorentz but not PT invariant, when one applies to it the standard quantization prescriptions within

[4]Bell himself requires that the equations be 'rational and integral', rather than polynomial, but it is unclear what this is supposed to mean.

the Lagrangian formalism. The above considerations suggest that the resulting quantum field theory will be 'pathological' in some sense, but it is not immediately clear in precisely which sense.

Relationship between the classical PT theorem proved here and the QFT CPT theorem.

As mentioned in the introduction to this chapter, when one starts with a classical theory and 'quantizes' it, one *ordinarily* expects that the process of quantization will not break the symmetry. This suggests the possibility of giving a new proof of the quantum-theoretic CPT theorem, by starting with the classical result and *proving* that, in the present case, quantization does not break the classical symmetry.

While this is an intriguing idea, the prospects of implementing it are, as far as I can see, rather dim. The connection between symmetries of a quantum theory and symmetries of its $\hbar \to 0$ 'classical limit' are clearest in the sum-over-histories formalism. (For an exposition of the sum-over-histories formalism for quantum field theory, see, for example, (Peskin & Schroeder, 1995, chapter 9).) However, in that formalism, to obtain a sufficient condition for some transformation on classical field configurations to be a symmetry of the theory, we must show not only that the transformation in question is a symmetry of the classical Lagrangian, but also that it is a symmetry of the functional measure. The latter task *seems* to be hopeless, since there is no obvious analog for functional measures of the (indispensable) condition that the Lagrangian be 'polynomial' (or similar).

3.5 Spinor field theories

It is noteworthy that the classical theorem stated and proved in section 3.3 applies only to systems of *tensor*-valued classical fields. Quantumly, of course, we have a CPT theorem that applies both to tensor-valued *and* to spinor-valued fields. This raises the question of whether or not there is also a classical 'PT theorem' for spinor field theories.

The answer — surprisingly — appears to be negative. This can be seen in two ways: by considering the mathematics of the proof of the tensor theorem, and by writing down an explicit counterexample to the analogous claim for the spinor case.

Consider, first, the proof given for the tensor case in section 3.3. The mathematical core of this proos was the fact that any polynomial function on the Lorentz group that vanishes on $L_+^\uparrow$ also vanishes on $L_+^\downarrow$. The analogous statement for spinors (it turns out) would be: any polynomial function on $SL(2,\mathbb{C}) \times SL(2,\mathbb{C})$ that vanishes on the subgroup consisting of elements of the form (A, A) also vanishes on elements of the form $(A, -A)$. (This is explained in, for example, (Greenberg, 2003, section 2).) But this latter statement is manifestly false: a counterexample is the polynomial $H(A, B) = A^0{}_0 - B^0{}_0$.

For our second reason for thinking that there cannot be a classical PT theorem for spinors, consider the following Lagrangian (intended as a counterexample to the claim that all *spinor* field theories that are $L_+^\uparrow$-invariant are also 'PT-invariant') for 'Dirac spinors':

$$\mathcal{L}(\psi) = \overline{\psi}\psi + \eta_{\mu\nu} \left(\overline{\psi}\gamma^\mu\psi\right) \left(\overline{\psi}\gamma^\nu\psi\right). \tag{3.14}$$

A 'Dirac spinor' is an element of the space $W \oplus \overline{W}^*$, where W is a two-dimensional complex vector space, and $\overline{W}^*$ is its complex-conjugate dual space. (For an exposition of the body of theory surrounding this statement, see, for example, (Wald, 1984, section 13.1).) The 'PT' transformation on classical Dirac spinors is, presumably[5], $\psi \mapsto i\gamma^5\psi$, where γ^5 is the block-diagonal matrix

$$\gamma^5 \equiv \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

and ψ is a column vector whose first two components represent an element of W and whose last two entries represent an element of $\overline{W}^*$.

Now, let us consider the following two 'Dirac bilinears': the 'scalar blinear' $\overline{\psi}\psi$ and the 'vector bilinear' $\overline{\psi}\gamma^\mu\psi$. (Here, $\overline{\psi} := \psi^\dagger\gamma^0$, and the γ^μ ($\mu = 0, 1, 2, 3$) are the 'Dirac gamma matrices', given (in block form) by

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$
$$\gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix},$$

$i = 1, 2, 3.$) Under PT, these quantities transform as follows:

$$\begin{aligned}
&\overline{\psi}\psi \\
\mapsto &(i\gamma^5\psi)^\dagger\gamma^0(i\gamma^5\psi) \\
= &\psi^\dagger(\gamma^5)^\dagger\gamma^0\gamma^5\psi \\
= &\psi^\dagger\gamma^5\gamma^0\gamma^5\psi && \text{since } (\gamma^5)^\dagger = \gamma^5 \\
= &-\psi^\dagger\gamma^0(\gamma^5)^2\psi && \text{since } \{\gamma^0, \gamma^5\} = 0 \\
= &-\psi^\dagger\gamma^0\psi && \text{since } (\gamma^5)^2 = 1 \\
= &-\overline{\psi}\psi;
\end{aligned}$$

[5] Our only argument in support of this presumption is that it is difficult to see what else it could be. Bell, for example, takes the quantum 'CPT' transformation to be $\psi \mapsto i\gamma^5\tilde{\psi}$, where $\tilde{}$ denotes transposition of the matrices representing operators on the quantum Hilbert space, and the classical analogs of such matrices are scalars.

and

$$\begin{aligned}
&\overline{\psi}\gamma^\mu\psi \\
\mapsto&(i\gamma^5\psi)^\dagger\gamma^0\gamma^\mu(i\gamma^5\psi) \\
=&\psi^\dagger(\gamma^5)^\dagger\gamma^0\gamma^\mu\gamma^5\psi \\
=&\psi^\dagger\gamma^5\gamma^0\gamma^\mu\gamma^5\psi \\
=&-\psi^\dagger\gamma^0\gamma^5\gamma^\mu\gamma^5\psi \\
=&+\overline{\psi}\gamma^\mu(\gamma^5)^2\psi && \text{since } \{\gamma^5,\gamma^\mu\}=0 \\
=&\overline{\psi}\gamma^\mu\psi.
\end{aligned}$$

The crucial thing here is that, whereas $\overline{\psi}\psi$ (respectively, $\overline{\psi}\gamma^\mu\psi$) transforms like a scalar (respectively, a vector) under *restricted* Lorentz transformations, it picks up an additional minus sign relative to a true scalar (respectively, a true vector) under the PT transformation. This allows us to construct restricted-Lorentz-invariant, PT-non-invariant polynomials. An example is the Lagrangian above: under PT, the first term of this Lagrangian flips sign, while the second term is not invariant, and hence the Lagrangian as a whole is not invariant even up to a scalar factor.

Thus, subject only to our claim that the transformation $i\gamma^5\psi$ *is* correctly regarded as the classical analogue of the quantum CPT transformation, we have shown that there cannot be a classical analogue of the CPT theorem for spinors, by writing down a counterexample.

3.6 Conclusions

The CPT theorem is usually thought of as an essentially quantum-theoretic result. But this is not entirely accurate: there is a classical analogue of the CPT theorem *for systems of tensor fields.* This classical result does not, however, extend to systems of spinor fields.

Chapter 4

Towards a geometrical understanding of the CPT theorem

4.1 Introduction

A story has it that in the early sixties, Feynman was asked to give an evening talk to physics students at Caltech, explaining the basic idea of the CPT theorem: the celebrated result in quantum field theory that states that any relativistic (i.e. Lorentz-invariant) quantum field theory must be invariant under CPT, the composition of charge conjugation, parity reversal and time reversal. Feynman agreed to commit to doing this, commenting that if one cannot explain something to second year Caltech undergraduates then one does not understand it. The story goes that Feynman spent a month or two trying to plan the talk, and then, in despair, cancelled the commitment.

Whether or not this story is true, its basic point is well taken: despite the importance of the CPT theorem in particle physics, the result itself is generally not well understood, even by those whose professional practice regularly appeals to it. It is often referred to as a 'remarkable result'. It seems worthwhile trying to attain a point of view from which the CPT theorem is not remarkable at all, but is, rather, precisely what one expects on elementary grounds. That is the aim of the project of which the present paper is a part.

More precisely, one can identify two positive sources of puzzlement:

- How can it come about that one symmetry (e.g. Lorentz invariance) entails another (e.g. CPT) *at all*?

- How can there be such an intimate relationship between *spatiotemporal* symmetries (Lorentz invariance, parity reversal, time reversal) on the one hand, and *charge conjugation*, not obviously a spatiotemporal notion at all, on the other?

This paper focusses on the first sort of puzzlement. I first sharpen the puzzle by suggesting that, according to a way of thinking about spacetime symmetries that is (for good reason) fairly common currency in the philosophy of physics community, there is a particular reason for thinking that Lorentz covariance should *not* be able to entail anything like CPT covariance. I then go on to offer a solution to the puzzle.

An outline of the paper is as follows.

Section 4.2 reviews the standard way of thinking about spacetime symmetries, well discussed by (in particular) Michael Friedman (1983) and John Earman (1989), that will give rise to the sharpened form of our puzzle and that will provide the framework for our discussion. The key point to be taken from this section, for the purposes of this paper, is that one generally expects to find a certain correspondence between the dynamical symmetries of a give spacetime theory, on the one hand, and the spacetime structure postulated by that theory, on the other. More precisely, we expect the following principle to hold: that the covariance group of a theory (when formulated non-generally-covariantly) should, for a well-formulated theory, be equal to the invariance group of the set of geometrical objects that are not represented explicitly in the coordinate-dependent, non-generally-covariant formulation in question. (Readers familiar with the standard framework in question can easily skim this section.)

Section 4.3 suggests that, from this point of view, the existence of a CPT theorem is *prima facie* puzzling. The idea here will be that a CPT theorem seems to be telling us that it is not possible for a relativistic theory (that is, on our way of thinking, a theory that does not require the existence of a preferred frame or foliation) to make essential use of a temporal orientation. Since a manifold with only a Lorentzian metric can be temporally orientable — capable of admitting a temporal orientation — this seems to be an odd sort of necessary connection between distinct existences; since there is no obstacle to theories' making essential use of *other* pieces of spacetime structure, such as a metric or a total orientation, we require an account of what makes temporal orientation special.

To anchor our discussion and to enable us to carry out its remainder in the simpler context of classical, rather than quantum, field theory, section 4.4 recalls the classical PT theorem proved in chapter 3, and discusses ways of formulating Lorentz-invariant, PT-violating theories by violating one or more of the 'auxiliary constraints' required for that theorem. This discussion shows that (as we would expect) Lorentz-invariant, PT-violating theories are not ruled out *as a matter of logical or mathematical consistency*. However, at this stage we will still have a puzzle about how and why the 'auxiliary constraints' suffice for the result.

Section 4.5 offers a solution to the puzzle: the key point is that temporal orientation is indeed (in a Lorentzian context) unlike many other pieces of spacetime structure, in that it cannot be represented by a Lorentz-invariant tensor field. Meanwhile, the 'auxiliary constraints' that we expect any 'reasonable' field theory to satisfy have the effect that only pieces of spacetime structure that can be represented by such tensor fields can be made use of in the theory.

Section 4.6 considers the puzzle and its suggested resolution in the context of *Galilean-* (rather than Lorentz-) invariant field theories. The point here is that while we do have a 'Lorentzian CPT theorem' — a theorem stating that any *Lorentz*-invariant field theory must also be CPT invariant — we do not have a 'Galilean CPT theorem' (and there do exist Galilean-invariant, CPT-non-invariant theories). We can therefore perform a 'sanity check' on the discussion of this paper, by checking that the suggested explanation of how anything like a CPT theorem can come about in the Lorentzian case does not also suggest that we should expect to find a CPT theorem in the Galilean case. The result will be reassuring: there *is* a *Galilean*-invariant tensor field representing temporal orientation, and, indeed, by making use of the field in question we can easily construct examples of Galilean-invariant, non-CPT-invariant field theories.

Section 4.7 is the conclusion.

4.2 The connection between dynamical symmetries and spacetime structure

This section reviews a standard way of thinking about spacetime symmetries. This standard account provides the framework within which the existence of the CPT theorem is, I will suggest, *prima facie* puzzling. The review in this section is very much in the spirit of the discussions given by Friedman ((1983), chapters 2 and 3) and Earman ((1989), chapters 2 and 3). It may be skimmed by those familiar with the framework in question (the only slightly idiosyncratic elements are the talk of 'special' rather than 'absolute' or 'kinematical' objects, and (relatedly) the terminology 'covariance$_Q$ group of a theory'; I indulge in this idiosyncrasy to avoid irrelevant complications concerning how, if at all, one might define 'absolute' or 'kinematical').

Spacetime theories.

Let T be a spacetime theory. That is, T is a theory whose intended models are structures of the form $\langle M, \Phi_1, \ldots, \Phi_n \rangle$, where M is a differentiable manifold, and the Φ_i are geometrical objects on M.

Let us suppose that these structures explicitly represent all the structure that is presupposed by the theory — for example, if the theory is supposed to be set in Minkowski spacetime, then one of the Φ_i's will be the Minkowski metric g. (We will formalize this condition below.)

Symmetries.

To discuss the symmetries of a theory T, we need first to regard the set M_D of models of the theory (D for 'dynamically allowed') as a subset, $M_D \subset M_K$, of a larger set M_K of 'kinematically allowed structures'. We then consider maps from the set M_K into itself; we say that such a map is a *symmetry* of the theory T iff the map leaves the dynamically allowed subset M_D invariant.[1]

Spacetime symmetries.

We wish formally to capture the sense in which particular groups of spacetime transformations — for example, the Lorentz or Galilei groups — may or may not be symmetries of a given spacetime theory T.

An abstract group is not itself a map from M_K into itself, so cannot literally *be* a symmetry of a theory in the sense just defined. Instead: an *action* A of a group G on a set M_K is a map from G to the set of injections from M_K into itself, respecting the group product operation (in the sense that for all $g_1, g_2 \in G$ and for all $m \in M_K$, $(A(g_1 g_2))(m) = (A(g_1))(A(g_2)(m)))$. We then say (speaking slightly loosely) that the elements of G are symmetries of the theory iff for all $g \in G$, $A(g)$ is a symmetry (where A is some particular action of G on M_K that we have in mind in making the statement).

Our particular interest is in *spacetime* symmetries. For any manifold M, we have the group *Diff*(M), and its defining action as a group of diffeomorphisms of M. We take it that, for any geometrical object Φ on a manifold M, there is a 'natural' action of the diffeomorphism group *Diff*(M) on Φ: for example, if Φ_i is a vector field on M, the natural action of h takes Φ_i to its push-forward $h_* \Phi_i$, while if Φ_i is a one-form then the natural action is the pull-back to $h^* \Phi_i$. (In the general case, the specification of this 'natural' action of the diffeomorphism group on Φ may be taken as part of the 'definition' of Φ. Thus, for example, true tensors and

[1]Note that this is a very minimal sense of 'symmetry', according to which any theory with N models has as many distinct symmetries as there are bijections from an N-element set onto itself). Most symmetries in this sense, of course, will be uninteresting; we will be interested only in those that are 'generated' in some particularly simple way.

so-called 'pseudo-tensors' are regarded as distinct types of geometrical object.)[2] In discussing spacetime symmetries, we are interested in finding subgroups G of *Diff*(M) with the feature that relative to the natural action, all elements of G are symmetries of the theory. There is a familiar obstacle to rendering such talk of spacetime symmetries *nontrivial*: to finding a sense, that is, in which it is not trivially the case that *any* manifold diffeomorphism $h : M \to M$ will count as a 'symmetry of T'. The obstacle arises from the fact that, in the structures $\langle M, \Phi_1, \ldots, \Phi_n \rangle$, we have explicitly represented all of the structure that is actually presupposed by our theory. It arises as follows. Write $h * \Phi_j$ for the result of allowing h to act in the natural way on Φ_j. Now, we wish to associate, with a manifold diffeomorphism $h \in$ *Diff*(M), a map $h' : M_K \to M_K$. The most obvious way to do this is to allow h to have its natural effect on *each* of the geometrical objects in an arbitrary structure $m \in M_K$: that is, to define $h' : M_K \to M_K$ by

$$h' \langle M, \Phi_1, \ldots, \Phi_n \rangle := \langle M, h * \Phi_1, \ldots, h * \Phi_n \rangle. \tag{4.1}$$

However, if we define h' in this way, then *every* h' (i.e. the h' corresponding to *every* $h \in$ *Diff*(M)) will be a symmetry of our theory T. This is the sense in which, unless there is some structure to M that we have failed to represent in our statement of the models of the theory, any spacetime theory will, trivially, be invariant under all of *Diff*(M). (The condition that M_D be diffeomorphism-invariant in this sense is thus the promised formal expression of our assumption that the structures in question 'explicitly represent all the structure that is presupposed by the theory'.)

---

[2] This appeal to 'natural' actions is supposed, in particular, to rule out 'deviant' actions according to which, say, one element of the group and its inverse each permute two otherwise unrelated elements of M_D while leaving the rest of M_K invariant, and all other group elements act trivially on M_K: we do not wish to recognize a sense in which (say) elements of the Galilean group are symmetries of Maxwellian electrodynamics simply on the grounds that group actions like *this* exist.

This is unhelpful, since we want to capture the special relationship of, say, the Lorentz group to relativistic electromagnetic theory, and the Galilei group to Newtonian gravitation theory. (One normally wants to say that Newtonian gravitation theory is Galilean-covariant and that Maxwell's equations are not, or that a theory counts as special relativistic just in case it is Lorentz-covariant; we seem to be losing an interesting and fruitful distinction if we have only the sense in which all theories are generally covariant.) To do this, we must set up a *different* correspondence between manifold diffeomorphisms h and maps from M_K onto itself, such that in general only for some *proper subset* of $\mathit{Diff}(M)$ do we have the corresponding maps as symmetries. Our new correspondence is set up as follows. For a given theory T, we single out some subset Q of the Φ_i as 'special'. (One way to go about branding objects 'special' is to look for some formal criterion that will pick some of them out, such as the Anderson-Friedman 'absoluteness' criterion (see, e.g., Friedman ((1983), pp.56-61). Another is to say that the 'special' ones are the 'kinematical' or 'geometrical' ones, and hope that we know what this means. An approach that is less ambitious, but that suffices for our present purposes, is to do without any such general criterion, and simply to specify some subset of the objects in a given theory on a case-by-case basis, putting a subscript on 'covariance' to indicate which set of objects we have chosen to treat as 'special'. Since we don't need to tangle with the problems that the more ambitious programs face, we will take this last approach.) Having chosen our set Q, we then write candidate models of T in the form $\langle M, S_1, \ldots, S_m, O_1, \ldots, O_n \rangle$, where the S_i ('special' objects) are elements of Q and the O_i ('ordinary' objects) are not. We now allow the diffeomorphism h to act only on the 'ordinary' objects $O_i \notin Q$. That is, to any $h \in \mathit{Diff}(M)$ we associate a map $h_Q : M_K \rightarrow M_K$, defined as follows:

$$h_Q\langle M, S_1, \ldots, S_m, O_1, \ldots, O_n\rangle := \langle M, S_1, \ldots, S_m, h * O_1, \ldots, h * O_n\rangle. \qquad (4.2)$$

We define the *covariance*$_Q$ *group of* T to be the set of $h \in \mathit{Diff}(M)$ such that the corresponding h_Q is a symmetry of T.

The connection between symmetries and spacetime structure.

The covariance$_Q$ group of a spacetime theory will, in general, be some proper subset of *Diff*(M). But more can be said. Define the *invariance group* of a set Q of geometrical objects as the set of diffeomorphisms h such that (the natural action of) h leaves each element of Q invariant. Suppose it is the case that, for all models of our theory T, the invariance group of the set $(S_1, \ldots, S_m)$ of 'special' fields appearing in that model is the same. In this case, we can write of the invariance group of Q as a property of the theory, rather than of a particular model of the theory. We then expect that, if our theory T is 'well-formulated', *the covariance*$_Q$ *group of* T *is equal to the invariance group of* Q.

To support this expectation, we argue first that the invariance group of Q is a subgroup of the covariance$_Q$ group of T, and then that the covariance$_Q$ group of T is a subgroup of the invariance group of Q. (Similar arguments are given in Earman (*ibid.*, pp.46-7).)

The first claim — that the invariance group of Q is a subgroup of the covariance$_Q$ group of T — follows trivially from the sense in which T is generally covariant. (Since

$$\begin{aligned} &\langle M, S_1, \ldots, S_m, O_1, \ldots, O_n\rangle \in M_D \\ \Rightarrow &\langle M, h * S_1, \ldots, h * S_m, h * O_1, \ldots, h * O_n\rangle \in M_D, \end{aligned}$$

if in addition we have $h * S_1 = S_1, \ldots, h * S_m = S_m$, it follows trivially that

$$\begin{aligned} &\langle M, S_1, \ldots, S_m, O_1, \ldots, O_n \rangle \in M_D \\ \Rightarrow &\langle M, S_1, \ldots, S_m, h * O_1, \ldots, h * O_n \rangle \in M_D; \end{aligned}$$

that is, that h_Q takes models to models.)

The second claim — that the covariance$_Q$ group of T is a subgroup of the invariance group of Q — can arguably be defended, for suitable selections of the set Q, by an appeal to Ockham's Razor. Here it is important that the ('special') objects in Q are not themselves 'directly observable' or 'given to us by a mechanical experiment': that their existence is, rather, inferred from empirical data that more directly gives us the 'ordinary' objects O_i. The basic idea is that, if we have a theory and a set Q of 'special' objects such that the invariance group of Q is a *proper* subset of the covariance$_Q$ group of T, then it ought to be possible to write down an alternative theory T' that has the same empirical consequences as does T as far as the O_i are concerned, but that replaces Q with a set Q' whose invariance group is larger than that of Q; further, that this alternative theory T' is more parsimonious than T. The claim then is that, *if T is a 'well-formulated' theory* (i.e. if T respects Ockham's Razor), the invariance group of Q will be a subgroup of the covariance$_Q$ group of T.[3]

[3]In theories that are genuinely generally covariant, such as general relativity (GR), the natural move is to take the set Q of 'special' objects to be the null set, in which case it is vacuously true that the invariance group of Q is *Diff*(M); that the covariance$_Q$ group of the theory is then also *Diff*(M) follows from the 'cheap' sense in which T is diffeomorphism-invariant. In this sense, the present account is consistent with the received wisdom that there is a *non*-trivial sense in which GR (but not SR) is diffeomorphism-invariant. However, one can then ask in virtue of what it is 'natural' to take Q to be null in such theories. The project of answering this question is of a piece with the (post-Friedman) project of supplying a criterion of 'absoluteness' according to which general relativity has no absolute objects, or otherwise precisely characterizing the sense in which general relativity is special. This project lies outside the scope of the present paper.

Non-generally covariant formulations of spacetime theories.

While it is often preferable, for the purposes of foundational discussions, to formulate theories in a coordinate-free framework, such a framework is often inconvenient for calculations, and is used in only a minority of the the physics literature. It will therefore be useful to see how the abstract considerations above relate to coordinate-dependent formulations of theories.

When formulating one's theory in a coordinate-dependent way, one faces a choice between two options. (The distinction between the two is precisely analogous to the distinction between the candidate symmetry operations h' and h_Q given above.) The first option is explicitly to take coordinate components of *all* the geometrical objects that appear in the coordinate-independent formulation. If one takes this first option, one arrives at a coordinate-dependent formulation that picks out the intended class of models relative to an arbitrary coordinate system. Say that the *covariance group* of the theory is the group of transformations between coordinate systems that pick out the intended class of models; we thus have, in this first case, $\mathit{Diff}(M)$ as the covariance group. The second, alternative, option is to represent some chosen subset Q of one's geometrical objects *implicitly*: that is, to consider its coordinate components as functions of the coordinates, and to 'transform' them, when changing to any other coordinate system, by keeping the same function of the coordinates in the new frame. If one takes this second option, one arrives at a coordinate-dependent formulation that picks out the intended class of models only relative to a certain 'privileged' class of coordinate systems (the 'privileged' class being the class of coordinate systems in which the coordinate components of the implicit geometrical objects happen to be the same as their components in the original, defining, coordinate system); its covariance group will then, in general, be some proper subgroup of $\mathit{Diff}(M)$, and again we expect that the covariance group will be equal to the invariance group of the set Q of objects that we chose to single out for special treatment.

Example.

We illustrate the above abstract discussion using the example of special-relativistic electromagnetism. According to this theory, there is a flat Lorentzian metric g_{ab}, a tensor field F_{ab} (the electromagnetic field) of type $(0,2)$, and a vector field J^a (the charge-current density field). The equations relating these objects are

$$\begin{aligned} F^{ab}{}_{;b} &= -4\pi J^a, && (4.3) \\ F_{[ab;c]} &= 0, && (4.4) \end{aligned}$$

where indices are raised using the inverse g^{ab} of the metric, and it is understood that the covariant derivative is the unique one that is compatible with the metric. *These* equations are generally covariant, in the following two (equivalent) senses:

Coordinate-independent sense of general covariance. If $\langle M, g_{ab}, F_{ab}, J^a \rangle$ satisfies (4.3) and (4.4), then so does $\langle M, h * g_{ab}, h * F_{ab}, h * J^a \rangle$, for any manifold diffeomorphism $h : M \to M$.

Coordinate-dependent sense of general covariance. In coordinate component form, the equations 4.3–4.4 become

$$\begin{aligned} F_{\mu\nu;\nu} &\equiv \frac{\partial F_{\mu\nu}}{\partial x^\mu} - \Gamma^\lambda{}_{\mu\nu} F_{\lambda\nu} - \Gamma^\lambda_{\nu\nu} F_{\mu\lambda} && (4.5) \\ &= J_\mu; && (4.6) \\ F_{[\mu\nu;\sigma]} &\equiv \frac{1}{3}(\frac{\partial F_{\mu\nu}}{\partial x^\sigma} - \Gamma^\lambda{}_{\mu\sigma} F_{\lambda\nu} - \Gamma^\lambda{}_{\nu\sigma} F_{\mu\lambda} && (4.7) \\ &\quad + \frac{\partial F_{\nu\sigma}}{\partial x^\mu} - \Gamma^\lambda_{\nu\mu} F_{\lambda\sigma} - \Gamma^\lambda_{\sigma\mu} F_{\nu\lambda} && (4.8) \\ &\quad + \frac{\partial F_{\sigma\mu}}{\partial x^\nu} - \Gamma^\lambda_{\sigma\nu} F_{\lambda\mu} - \Gamma^\lambda_{\mu\nu} F_{\sigma\lambda}) && (4.9) \\ &= 0. && (4.10) \end{aligned}$$

> These equations pick out the same (i.e. the intended) class of models in *any* coordinate system $x : M \to \mathbb{R}^4$.

However, we can also identify a clear sense in which 'the symmetry group of classical electromagnetism' is the Lorentz group, rather than the full diffeomorphism group:

Coordinate-independent sense of special covariance. Let us single out the metric g as 'special'. Then, for any $h \in \mathit{Diff}(M)$, we may consider the transformation h_g, given by

$$h_g\langle M, g_{ab}, F_{ab}, J^a\rangle := \langle M, g_{ab}, h * F_{ab}, h * J^a\rangle. \tag{4.11}$$

For *arbitrary* h, we won't in general expect *this* transformation to take models to models. In general we'll (instead) expect h-covariance$_g$ only when h happens to leave g invariant, since, in that case but in that case alone, the RHS of 4.11 is identical to $\langle M, h * g_{ab}, h * F_{ab}, h * J^a\rangle$. So now we have a nontrivial covariance$_g$ group, and it's precisely the group of transformations leaving the 'special' object g invariant: that is, the Lorentz group.

Coordinate-dependent sense of special covariance. If we choose a coordinate system in which the Christoffel symbols vanish (i.e. an inertial coordinate system), then, the equations (4.5)–(4.10) reduce, respectively, to

$$\frac{\partial F_{\mu\nu}}{\partial x^\nu} = J_\mu; \tag{4.12}$$

$$\frac{\partial F_{\mu\nu}}{\partial x^\sigma} + \frac{\partial F_{\nu\sigma}}{\partial x^\mu} + \frac{\partial F_{\sigma\mu}}{\partial x^\nu} = 0. \tag{4.13}$$

(Noting that

$$F = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_x \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}, \tag{4.14}$$

it is straightforward to see that these coincide with usual coordinate-dependent form of the Maxwell equations.)

We have gained notational simplicity, relative to (4.5)–(4.10), but now we must remember that our equations (4.12)–(4.13) pick out the intended class of models only relative to a privileged class of coordinate systems, which latter are related to one another by Lorentz transformations.

This concludes our review of the standard material within which our puzzle will appear. To sum up the key point of this section: there is an intimate relationship between the spacetime symmetries of a theory, on the one hand, and the spacetime structure postulated by that theory, on the other. Specifically, a 'well-formulated' theory fails to have a particular manifold diffeomorphism as one of its symmetries iff it postulates some piece of background ('special') structure that is not invariant under the diffeomorphism in question.

4.3 A puzzle about the CPT theorem.

We are now in a position to state our puzzle concerning the CPT theorem. This theorem states that, subject to some apparently innocuous auxiliary conditions, the following conditional must hold of any quantum field theory T:

If T is invariant under the restricted Lorentz group $L_+^\uparrow$, then T is actually invariant under CPT.

I mentioned (in the introduction) that it is possible to decompose a general sense of puzzlement at this statement into two parts: one concerning how Lorentz invariance can entail another symmetry at all, and a second concerning how charge conjugation gets into an otherwise spatiotemporal picture. Since our present concern is with the first of these, let us 'pretend' (but justification for this move will be offered in the next section) that, instead of the CPT theorem, we actually a PT theorem. Then we have (instead) the statement

> If T is invariant under the restricted Lorentz group $L_+^\uparrow$, then T is actually invariant under the whole of the proper Lorentz group L_+ (i.e. under the total-reflection component, as well as under the identity component).

In the light of the standard account of spacetime symmetries that I've reviewed, this conditional is *prima facie* rather puzzling. Here is why. Suppose that we have a theory according to which there are, among other objects, a flat Lorentzian metric g, a total orientation ϵ and a temporal orientation τ. (The total orientation is an object that determines, for any quadruple consisting of one timelike and three (ordered) linearly independent spacelike 4-vectors, whether that quadruple is 'right-handed' or 'left-handed'. It can be represented by a totally antisymmetric rank four tensor, ϵ_{abcd}. The temporal orientation is an object that specifies in a continuous way, at each point p, which is the 'future' lobe of the lightcone in T_pM. *Its* possible representations will be considered in section 4.5.)

The puzzle is then the following. First, we note the invariance groups of three sets of objects we might choose to treat as 'special' in the sense of section 4.2:

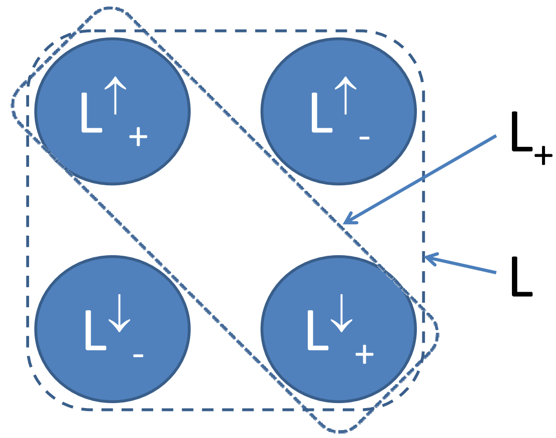


Figure 4.1: The (real) Lorentz group has four mutually disconnected components, labelled by $\uparrow$ or $\downarrow$ according to whether or not they reverse time sense, and by $+$ or $-$ according to whether or not they reverse total orientation (i.e. whether their determinant is $+1$ or -1). In this notation, $L_+^\uparrow$ is the 'restricted' Lorentz group; $L_+ \equiv L_+^\uparrow \cup L_+^\downarrow$ is the 'proper' Lorentz group; $L \equiv L_+^\uparrow \cup L_-^\uparrow \cup L_-^\downarrow \cup L_+^\downarrow$ is the 'full' Lorentz group.

Special fields	**Invariance group $\mathbf{S_k}$**
g	L (full Lorentz group)
g, ϵ	L_+ (proper Lorentz group)
g, ϵ, τ	$L_+^\uparrow$ (restricted Lorentz group)

Here, $L_+^\uparrow$ is the *restricted* Lorentz group. This is the set of Lorentz (i.e. g-preserving) transformations that can be continuously connected to the identity: it includes all rotations, boosts and products thereof, but does not include parity or time reflection. L_+ is the *proper* Lorentz group: the set of all metric-preserving Lorentz transformations with determinant one, i.e. the union of $L_+^\uparrow$ with the set of all Lorentz transformations that reverse *both* spatial parity and time sense. L is the *full* Lorentz group: this includes transformations that reverse parity, time sense, both or neither. (See figure 1.)

Ignoring the first of the possibilities listed in the above table (i.e. that of treating g alone as 'special'), we should then expect to be able to write down, not only a non-generally invariant theory whose invariance group is exactly L_+ (by treating g and ϵ as 'special'), but also a non-generally invariant theory whose invariance group is exactly $L_+^\uparrow$ (by treating g, ϵ and τ as 'special'). A PT theorem, however, tells us that we cannot do the latter: that, subject to the (as yet unstated) auxiliary assumptions of our theorem, we cannot find theories that are invariant under precisely the restricted Lorentz group. It seems to be telling us, that is, that no theory that is 'nice' (in the sense of conforming to these auxiliary assumptions) can actually make use of a temporal orientation, over and above a flat metric and a total orientation. And now one might well wonder *why not*. Metric, temporal orientation and total orientation seem to be paradigm cases of distinct existences; it's odd to find such necessary connections between them. Or, to put the puzzle another way: where does this discrimination against temporal orientations come from? That is, what feature of temporal orientation can explain why, in the context of the existing objects g and ϵ, they are unusable in this way?

This is not a paradox, but it does seem to be a puzzle whose resolution is likely to be illuminating. In the next section, I give an explicit statement of the theorem that is the source of our puzzle, and in section 4.5 I offer a resolution.

4.4 A classical PT theorem

At the start of section 4.3, I promised some motivation for changing the subject from CPT to PT. Here, briefly, are three reasons for going along with that. (The first is pragmatic; the second and third are more justificatory.)

1. The account of the connection between symmetries and 'special fields' has been developed only for *spacetime* symmetries: transformations that are

diffeomorphisms on the spacetime manifold M. So PT is a relatively well-understood case to deal with. It's less clear how we're supposed to think about CPT-invariance (i.e. whether, and how, this 'combination of a geometrical and a nongeometrical symmetry' can be associated with the absence of some piece of structure on some larger space). As a research strategy, to avoid unmanageable confusion at the outset, it seems worth starting with the more straightforward case of PT, and hoping that the results will generalize to CPT.

2. I have growing suspicions that the transformation usually called 'CPT' is actually more properly regarded as PT, that is, as a *bona fide* spacetime symmetry that is the product of mirror-image reflection and time reversal. If this is right, then I am not actually changing the subject at all, in insisting on talking about PT. But that's another story.[4]

3. Again to avoid irrelevant complications, it will be better to start the foundational discussion in the context of *classical* (rather than quantum) field theory. And in the classical case, we can prove a theorem that transparently *is* a PT, rather than a CPT, theorem. (Its precise relationship to the usual quantum 'CPT' theorem is an open question — but it seems inevitable that there will be an intimate connection, and this is part of my reason for thinking that so-called 'CPT' is really just PT.)

The following, then, is a theorem ('classical PT theorem').

[4]This comment is related to the idea that the unorthodox 'Feynman' time reversal operation considered in chapter 2 is, from the geometrical point of view, better deserving of the name 'time reversal' than is the standard or 'Malament' time reversal operation.

4.4.1 Bell's theorem

The theorem in question was stated and proved in chapter 3. We reproduce our informal summary of the content of the theorem here, for convenience; for the more rigorous version, see the preceding chapter.

> In outline, the result is as follows. ...We consider a classical theory given by a system of partial differential equations (PDEs) on a specified set of fields. Let Φ be the space of kinematically allowed fields. (In the general case, we may be dealing with a theory containing a number of interacting fields — scalar fields, tensor fields, etc — so, for a given theory, an element of Φ will be an ordered m-tuple of specified numbers of scalar fields, vector fields, rank 2 tensor fields, etc.) We note that any PDE can be expressed as the vanishing of some functional $F : \Phi \to \mathbb{R}^M$ of the fields. (That is, F encodes the dynamics in the sense that: $\phi \in \Phi$ is dynamically allowed iff $F(\phi)$ is the zero map on M.) We *assume that F is a 'local' polynomial in the fields and their derivatives.* It can then be proved that, if the set S of solutions of the equation $F(\phi) = 0$ is invariant under the tensor representation of the restricted Lorentz group $L_+^\uparrow$, then S is actually invariant under the tensor representation of the whole of the proper Lorentz group L_+ (i.e. including total reflections as well as rotations and boosts).

Summing this up, the claim is: Let T be a theory according to which there are n dynamical fields $\Phi_1, \ldots, \Phi_n$. Suppose that the following three conditions hold:

1. The dynamical fields are tensors (of arbitrary rank).
2. The dynamical equations are partial differential equations that are local polynomials in the fields and their derivatives.

3. The set $\mathcal{S}$ of solutions to the dynamical equations is invariant under $L_+^\uparrow$.

Then, $\mathcal{S}$ is actually invariant under all of L_+ $\left(\equiv L_+^\uparrow \cup PT(L_+^\uparrow)\right)$.

4.4.2 Auxiliary constraints

We were careful, in section 4.3, to state our puzzle as arising from the fact that no 'nice' theory is invariant under precisely the restricted Lorentz group, rather than that *no theory whatsoever* has just that invariance group. 'Nice', here, means 'conforming to the auxiliary assumptions of the PT theorem' (i.e. the conditions (1) and (2) above). It is worth highlighting, then, the fact that these 'innocuous auxiliary constraints' play a crucial role in both the antecedent plausibility, and in the proof, of the PT theorem. There obviously do exist 'theories', in the minimal sense of 'classes of models', that are $L_+^\uparrow$-invariant but not L_+-invariant. (To generate one, we need only pick some particular scalar field on $\langle M, g\rangle$ that does not have any interesting symmetries, and take the set that results from closing under the action of the restricted Lorentz group.)

But we would like to know more: we would like to be able to see precisely how it is that the particular auxiliary constraints in question — which, after all, *do* look pretty innocuous — manage to rule out the use of a temporal orientation.

In the classical theorem sketched above, our principal auxiliary constraint is a restriction on the dynamics: the dynamics must express the vanishing of all members of some particular set of polynomials in the coordinate components of the fields and their derivatives. There are two points here that are worthy of note. The first is that some equations, fairly plausible from the point of view of physics, are not 'polynomial' in the required sense.[5] It may be possible to weaken the assumptions of the theorem, so as to cover these cases also. The second point

---

[5]For example, equations involving terms like $\sin(\nabla_a\phi\nabla^a\phi)$ or operators like $\sqrt{\nabla^2+m^2}$ are not 'polynomial'. I am grateful to Robert Geroch and Michael Kiessling (resp.) for pointing out these particular examples.

is that some 'dynamics' do not express the vanishing of any mathematically simple functional F at all (polynomial or otherwise). One example of this phenomenon is given by the 'theory' sketched at the start of this subsection; another is given by the theory **'Inc'** stated below (see footnote 8).

It is also worth noting that by 'invariant under all of L_+, we mean: invariant under *the tensor representation of* L_+ (in the sense defined in chapter 3. The point here is that in the physics (as opposed to the mathematics) literature, one talks of 'how objects transform under PT' as part of the definition of those objects, rather than as part of the specification of which transformations on the set of fields one is making a claim about. In the physics-literature language, the restriction of our claim to tensor representations amounts to a substantive assumption about 'which types of fields' may be present in our theory: we are ruling out theories 'containing dynamical fields that transform as pseudotensors under PT'. If we are allowed PT-pseudotensors, then counterexamples to the claim of PT invariance are easy to come by. Here's one: let ϕ be a pseudoscalar field, and let the dynamical equation be

$$\phi = 1. \tag{4.15}$$

Slightly less trivially, suppose that ψ is a scalar field and χ a pseudoscalar under PT (i.e. $\chi \overset{e}{\longmapsto} \chi$ for $e \in L_+^\uparrow$, but $\chi \xrightarrow{PT} -\chi$). Then, the equation

$$\psi\chi - \psi = 0 \tag{4.16}$$

is $L_+^\uparrow$-invariant but not PT-invariant. (In mathematics-literature language: the equation is not invariant under 'PT-pseudotensor' representations of L_+.)

Be this as it may, there still seems to be something *prima facie* puzzling even about the restricted claim that all theories *within the stated class* obey the conditional 'if $L_+^\uparrow$-invariant then L_+-invariant' (relative to representations of L_+ within the stated class) — there is no connection yet apparent between the restrictions involved in the assumptions of the theorem on the one hand, and the surprising ineffectiveness of temporal orientation on the other. This is the puzzle we wish to solve.

4.5 Resolution of the puzzle

Let us take stock. We started (section 4.2) by sketching a way of thinking about spacetime symmetries according to which the set of dynamical symmetries ought to coincide with the invariance group of a set of objects that we have (for some reason or none) decided to single out as 'special'. We then noted (section 4.3) that, on this way of thinking, a PT theorem seems to be asserting that, subject to apparently innocuous auxiliary constraints, it is not possible to write down a theory that makes essential use of a temporal orientation, over and above a Lorentzian metric and a total orientation, and that this is puzzling. To ground the discussion, we then recalled (in section 4.4) an example of such a theorem, for the case of classical field theory. We now seek a more enlightened point of view: a point of view from which the existence of such theorems in certain cases is not puzzling at all, but is, rather, precisely to be expected, where and only where they in fact occur.

My suggestion is that the following observation lies at the heart of the otherwise puzzling nature of the CPT theorem: *there is no tensor field that represents temporal orientation and no more*, in the context of a flat Lorentzian metric and a total orientation.

The remainder of this section has two aims. The first is to explicate this observation — what exactly it means, and why it is true. The second is to explain how this helps to dissolve the puzzle. It will be easiest to tackle both of these aims simultaneously.

Intuitively, a temporal orientation on a (temporally orientable[6]) manifold M is supposed to specify which temporal direction is 'the future'. Let p be an arbitrary point in a temporally orientable manifold M that is equipped with a Lorentzian metric g. Then, the tangent space T_pM can be divided into timelike, spacelike and null vectors. Further, the set of *timelike* vectors in T_pM has two disconnected components: these will be the 'past' and 'future' lobes of the lightcone at p ('will be' rather than 'are', because until and unless we have a temporal orientation, neither lobe is distinguished as the 'future' one).

Now, we wish to represent temporal orientation by some geometric object on M. Here we have a choice: there are many structures on M that would do the trick.

The most obvious way (perhaps) of representing temporal orientation is by a map that assigns, to each point $p \in M$, one of the two lightcone lobes in T_pM (and that does so in a continuous way, i.e. the assignments of lightcone lobes to neighboring points must be 'mutually consistent'). This is our first candidate way of representing temporal orientation.

But let us now recall the use we wish to make of our pieces of spacetime structure: we wish to formulate laws that relate other ('dynamical'/'matter') fields to them, so that, by treating the spacetime structures as 'special', we can restrict the invariance groups of non-generally-invariant formulations of our theories. We then note that, if, as seems to be usually the case, our physical laws take the form of *differential equations* coupling various geometrical objects to one another, then a 'map from spacetime points to lightcone lobes' is not an object we can easily

[6]Definition: A manifold M equipped with a Lorentzian metric g is said to be *temporally orientable* iff there exists a continuous, nowhere-vanishing, timelike vector field on M. Heuristically: iff a manifold M fails to be temporally orientable, then one can 'parallel-transport' a timelike vector v at some point $p \in M$ around the manifold, and return to the point p with a vector v' that points in the opposite temporal direction to v. In this case, it is not possible to make any continuous global specification of which temporal direction is 'the future'.

work with. The point is that if f is such a map, the idea of a 'differential equation for f' does not seem to make sense; f, that is, is not the right sort of object to appear in differential equations.[7]

This observation suggests a second possible way of representing temporal orientation. Instead of using a map from spacetime points to lightcone lobes, we could use a *continuous nonvanishing timelike vector field*, t^a, on M. (We can then pick out the 'future'-directed timelike vectors $v^a \in T_pM$ as those that have positive 'dot product' $g_{ab}v^a t^a$ with t^a (relative to a convention according to which the metric has signature $(+,-,-,-)$ rather than $(-,+,+,+)$).) This move solves the problem we faced when trying to make use of f: t^a, as a vector field, *is* an object of a type that we perfectly well know how to use in differential equations. However, we have now incurred a problem of a different sort: t^a *is not restricted-Lorentz invariant*. That is, it is not the case that, $\forall l \in L_+^\uparrow, l * t^a = t^a$. The point here is that t^a picks out more structure than we wanted to pick out: we wanted only to pick out a preferred lobe of the lightcone at each point, but a vector field picks out, in addition, a preferred *timelike vector* in the chosen lightcone lobe. The upshot of this is that when we combine our 'temporal orientation' t^a with our existing pieces of structure g_{ab}, ϵ_{abcd}, we do not have a set whose invariance group includes $L_+^\uparrow$: rather, the most we expect is the group of translations and rotations (if g_{ab} is flat and v^a is constant).

This observation suggests a *third* possible way of representing temporal orientation: rather than a *single* (continuous nowhere-vanishing timelike) vector

[7]A referee for this paper pointed out that in a theory in which *discontinuous* fields are allowed (for example, a theory of a one-particle GRW wavefunction defined on spacetime), differential equations could use this first representation of temporal orientation in requiring the derivatives of the field *in one temporal orientation but not the other* to satisfy some equation. Such theories are ruled out by *fiat* in the conditions for the theorem stated in chapter 3 (it is assumed there that the fields be *tensor fields* — i.e., among other things, that they be smooth). We then face the questions: is there some other proof that Lorentz-invariance entails PT-invariance for 'discontinuous field theories' of this sort; if so, what prevents theories of *this* sort from making use of a temporal orientation; if not, can we write down a counterexample? These questions warrant further investigation.

field, we could take an *equivalence class* of such vector fields (where $s^a \sim t^a$ iff $g_{ab}s^a t^b > 0$). But now we are back to our original problem: an *equivalence class* of vector fields, as opposed to a particular vector field, is not the right sort of object to appear in a single partial differential equation.[8]

More generally: suppose we convince ourselves that the geometric objects we can make use of, in equations that satisfy the restrictions we have laid down, are just those that can be represented by tensor fields.[9] Then, we can avail ourselves of the following mathematical fact:

> There does not exist any Lorentz-invariant tensor field with an odd number of spacetime indices.[10]

---

[8]Such an equivalence class of vector fields *can*, of course, be used to generate a *set* of differential equations. Here is a non-PT-invariant theory that makes use of this idea: Take the temporal orientation τ to be the set of all nowhere vanishing, future-directed timelike vector fields. Let there be (besides the temporal orientation, total orientation and metric) a single scalar field ϕ. Say that ϕ is dynamically allowed iff the following condition holds:

(Inc) There exists at least one vector field $v^a \in \tau$ such that, at every spacetime point $p \in \mathbb{R}^4$, $v^a \nabla_a \phi > 0$.

(This theory is cooked up to say, in a restricted-Lorentz-invariant way, 'ϕ increases towards the future', and hence not to be PT-invariant.)

This example shows that the restrictions on the dynamics that appear in the premises of the theorem include restrictions on the 'logical form' of the dynamics: it's crucial to the theorem that the sort of existential quantification that's going on in this example is disallowed.

[9]It is not entirely clear that this is true. For example, the covariant derivative is usually thought of as a map from tensor fields of type (n, m) to tensor fields of type $(n, m+1)$, and not itself as a tensor field; and yet it can be used in PDEs. This suggests that *perhaps* the present discussion must be extended to some class of geometric objects that is wider than the class of tensor fields. However, it is also true that the covariant derivative *can* be represented by a tensor field (viz. the metric — since the covariant derivative operator is uniquely determined by the metric), so perhaps not.

A second point in this vein is that I am ignoring the issue of density weight. When one writes "tensor" rather than "tensor density of weight n", one normally implies that the object under discussion has density weight zero. I do *not* intend this implication. Density weight is irrelevant for present purposes, since all the transformations under consideration have determinant unity. (For an explanation of the concept of density weight, see, e.g., (Anderson, 1967), pp. 23-5.)

[10]Proof: We first note the existence of a type (0,2) and of a type (2,0) restricted-Lorentz-invariant tensor, namely, the Minkowski metric η_{ab} and its inverse η^{ab}. Second, we note that there does not exist a nonzero restricted-Lorentz-invariant tensor of type (0,1) or of type (1,0). (This latter is easy to prove by considering a coordinate system — which always exists — in which such a tensor has the form $(a, 0, 0, 0)$ — and noting that, since there is always a restricted Lorentz transformation that nontrivially mixes the first coordinate with the second, it cannot be that

It is easy to see, meanwhile, that a tensor representing the temporal orientation would *have* to have an odd number of spacetime indices; to represent temporal orientation, it must not be invariant under PT, but any tensor with an even number of spacetime indices *is* invariant under PT.

We can then conclude, as stated, that no Lorentz-invariant tensor can represent temporal orientation; given our conviction, it follows that no Lorentz-invariant tensor field theory can make use of a temporal orientation.

4.6 Galilean-invariant field theories

We now wish to perform a sanity check on the suggestion of section 4.5, by considering the case of *Galilean*-invariant field theories.

The point here is that the PT theorem does not hold in the Galilean case. That is, the following hypothesis is false:

Galilean PT hypothesis. If T is a spacetime theory containing tensor fields, whose dynamics are polynomial in the fields and their derivatives, and if in addition T is invariant under the restricted Galilean group $G_+^\uparrow$, then T is PT-invariant.

Therefore, if our suggested explanation of the possibility of a Lorentzian PT theorem is on the mark, it had better not be the case that the analogous statement is also true in the Galilean case. That is, it had better not also be true that there

---

such an object has the same coordinate representation in every nonreflected inertial coordinate system.)

Now, consider an arbitrary tensor $T^{a_1 \ldots a_n}{}_{b_1 \ldots b_m}$, with $n+m$ odd; we will show that T cannot be restricted-Lorentz-invariant. To show this, suppose WLOG that n is even, m odd. Define a new tensor T', of type $(0,1)$, as follows:

$$(T')_b := \eta_{a_1 a_2} \ldots \eta_{a_{n-1} a_n} \eta^{b_1 b_2} \ldots \eta^{b_{m-2} b_{m-1}} T^{a_1 \ldots a_n}{}_{b_1 \ldots b_{m-1} b}. \tag{4.17}$$

Then, if T were restricted-Lorentz-invariant, T' would be too. But this is impossible, since we have already shown that there are no rank 1 restricted-Lorentz-invariant tensors.

is no way of representing temporal orientation against a background of *Galilean* spacetime structure, without 'picking out more structure than we want', i.e., in this case, the object that represents temporal orientation had better be invariant under the restricted Galilean group.

At first sight, things look worrying. One of the points we met in the Lorentzian case was that a vector field picked out a timelike direction, as well as a privileged direction of time. But privileged timelike directions are no more acceptable in the Galilean setting than in the Lorentzian.

Fortunately for our suggested explanation, however, it does not, in fact, also go through in the Galilean case. There is no Galilean-covariant *vector* field, but there *is* a Galilean-covariant *one-form* (corresponding to the fact that, in Galilean spacetime, there is no preferred timelike direction, but there is a priveleged notion of simultaneity). In this section we explain this point, and we use it to develop a counterexample to the Galilean PT hypothesis.

4.6.1 Temporal orientation in Galilean spacetime

One encodes the structure of Minkowski spacetime using a flat Lorentzian metric g; elements of the (full) Lorentz group are then transformations leaving g invariant. Things are less simple in the Galilean case: there is no single geometric object, as it were, that will encode, in a single shot, all of the structure of Galilean spacetime.

Let us first get clear about what the structure is that we are trying to encode, over and above topological and differential structure. To model the Galilean case, we want our spacetime to possess a natural foliation into a family of three-dimensional hypersurfaces, the preferred simultaneity slices. We want each simultaneity slice to be equipped with a Euclidean spatial metric. We want there to be a fact, for any two points of spacetime, about what is the (absolute value of the) temporal distance between them. And we want there to be a fact, for any timelike curve, about whether or not it is 'straight' (i.e. is an inertial trajectory). *Iff* we want to endow our Galilean spacetime with a temporal orientation, then we also want there to be a privileged total ordering on the set of simultaneity slices.

One way of encoding the aspects of this structure, aside from temporal metric and temporal orientation, is as follows (here I largely follow Friedman ((1983), pp. 71-92), who sets this approach out in far more detail). We start, as in the Lorentzian case, with a four-dimensional differentiable manifold M. The affine structure (i.e. the set of facts about which lines in the spacetime are 'straight') is encoded by a connection Γ. The Euclidean spatial metrics are encoded by a rank 2 tensor field h^{ab}.

We now face the question of how to encode the temporal metric and/or temporal orientation. Suppose first that we wish to encode temporal metric *without* picking out a preferred temporal orientation. This can be done by means of a symmetric tensor field of type $(0, 2)$ (satisfying certain restrictions; cf. Earman (1989), pp.30-31; in Earman's notation, the tensor field in question is h_{ij}). This object will tell us the temporal distance between any two time-slices, but will not tell us which is to the future of which. Second, though, suppose that we do wish to encode a temporal orientation, in addition to a temporal metric. Then, we can use a one-form, t_a (this can be thought of as the exterior derivative, $t_a := (dt)_a$, of a global time function t that respects the simultaneity structure in the sense that the surfaces of constant t are the simultaneity surfaces). This represents temporal metric and temporal orientation at once, in the natural way: if v^a is a timelike vector, then $|t_a v^a|$ is the temporal length of that vector, and the sign of $t_a v^a$ tells us whether v^a is future- or past-directed. And t_a can be chosen to be invariant under the restricted Galilean group $G_+^\uparrow$, so we have not picked out more structure than we wished to encode.

4.6.2 Counterexample to the Galilean PT hypothesis

The above suggestion for encoding temporal orientation in Galilean spacetime, via the one-form t_a, can easily be used to generate a counterexample to the 'Galilean PT hypothesis' above. Here is one such counterexample:

> Suppose we have a theory containing a scalar field ϕ and vector field v^a, whose dynamics are given in generally covariant form by single equation
>
> $$t_a v^a = h^{ab}\phi_{;a;b}. \tag{4.18}$$
>
> Here, t_a, h^{ab} are understood as, respectively, the temporal structure and Euclidean spatial metric structure outlined above for Galilean spacetime.
>
> Suppose now that we treat t_a and h^{ab}, and in addition the flat connection Γ, as 'special'. Then, we have a privileged class of coordinate systems: the inertial frames in which t increases towards the future. In these coordinate systems, the dynamics is given by the non-generally-covariant equation
>
> $$v^0 = \nabla^2\phi. \tag{4.19}$$
>
> Under a restricted Galilean transformation, both v^0 and $\nabla^2\phi$ are invariant. However, under PT, $\nabla^2\phi$ is invariant while v^0 flips sign. Hence, PT in general does not take solutions to solutions, while restricted Galilean transformations do. So this theory constitutes a counterexample to the Galilean PT hypothesis.

4.7 Conclusions

The existence of a PT theorem (such as that discussed in this paper) is *prima facie* puzzling, since it seems to show that a reasonable theory cannot make use of a temporal orientation, over and above a flat Lorentzian metric and total orientation, without also using extra, 'unwanted' structure such as a preferred frame. One might well wonder where this discrimination against temporal orientations comes from. This paper has suggested that temporal orientation in a relativistic context indeed is special, as pieces of spacetime structure go: it cannot be represented by a tensor field. Meanwhile, we seem to be committed to constraining principles on our physical theories (for example, constraints on the types of PDEs theories may use), such that structure that cannot be encoded via tensor fields (or 'similar') cannot be used. This dissolves the puzzle.

The discussion above was carried out in the classical context, using a 'classical PT theorem'. However, the hope is that the same sort of line of thought can be used to illuminate the CPT theorem in quantum field theory. My conjecture is that, from the point of view of field theory (as opposed to particle phenomenology), the operation usually called 'CPT' is in fact more naturally regarded as a PT-reversing operation, so that the 'CPT' theorem is also, properly understood, a PT theorem; and furthermore that precisely analogous proofs can be given for the classical- and quantum-theoretic PT theorems. (Both these suggestions are also made by Bell (1955).) One then wonders whether similar lines of thought can also illuminate the relationship between parity violation and 'CP' violation, the possibility of CPT violation, and so on.

Several open questions remain. The most pressing is perhaps the following. We have restricted attention thus far to *tensor* field theories. But this is not, of course, the most general type of field theory of physical interest, or for which we have (quantum-mechanically) a CPT theorem. In particular, the above treatment has said nothing about *spinor* field theories. In fact, preliminary investigation suggests that the case of spinors is rather complicated: classically, one *can* construct a Lorentz-invariant temporal orientation from spinors; correspondingly, classical spinor field theories *can* be Lorentz- but not PT-invariant; but somehow this possibility seems (given that we have a quantum CPT theorem covering spinor theories!) to vanish in the classical-to-quantum transition. The further investigation of these matters is a topic for a future paper.

References

Albert, D. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Anderson, J. L. (1967). *Principles of relativity physics*. New York/London: Academic Press.

Bell, J. S. (1955). Time reversal in field theory. *Proceedings of the Royal Society of London A*, *231*(1187), 479-95.

Borchers, H. J. (2000, June). On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics*, *41*(6), 3604-3673.

Earman, J. (1989). *World enough and space-time*. Cambridge, MA: MIT Press.

Feynman, R. (1985). *QED*. Princeton: Princeton University Press.

Friedman, M. (1983). *Foundations of space-time theories*. Princeton: Princeton University Press.

Geroch, R. (1973). *Special topics in particle physics.* (Available online from http://home.uchicago.edu/∼seifert/geroch.notes/gpp.pdf.)

Greenberg, O. W. (2003). *Why is CPT fundamental?* (Available online from www.arxiv.org.)

Isham, C. J. (1999). *Modern differential geometry for physicists* (2nd ed.). Singapore: World Scientific.

Itzykson, C., & Zuber, J.-B. (2005). *Quantum field theory*. Dover.

Malament, D. (2004). On the time reversal invariance of classical electromagnetic theory. *Studies in History and Philosophy of Modern Physics*, *35B*(2), 295-315.

Peskin, M., & Schroeder, D. (1995). *Introduction to quantum field theory*. Westview Press.

Streater, R., & Wightman, A. (1964). *PCT, spin and statistics, and all that.* New York: W. A. Benjamin.

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

Vita

Hilary Greaves

Colleges attended

2003-2008 Ph.D. in Philosophy, Rutgers University

1998-2003 B.A. in Physics and Philosophy, University of Oxford

Positions held

2007– Junior Research Fellow, Merton College, University of Oxford

2003–2007 Graduate student, Department of Philosophy, Rutgers University

Publications

Everett and Evidence. With Wayne Myrvold. Forthcoming in the joint proceedings of conferences on the Everett interpretation held in Oxford in July 2007 and at Perimeter Institute in September 2007, edited by Jonathan Barrett, Adrian Kent, Simon Saunders and David Wallace.

On the Everettian epistemic problem (March 2007). In *Studies in History and Philosophy of Modern Physics* 38(1), pp. 120–152.

Probability in the Everett interpretation (January 2007). In *Philosophy Compass* 2(1), pp. 109–128.

Justifying conditionalization: Conditionalization maximizes expected epistemic utility (July 2006). With David Wallace. In *Mind* 115(459), pp. 607–632.

Understanding Deutsch's probability in a deterministic multiverse (September 2004). In *Studies in History and Philosophy of Modern Physics* 35(3), pp. 423–456.