Prospects for a New Account of Time Reversal

by

Daniel J. Peterson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2013

Doctoral Committee:

Professor Gordon Belot, Chair
Assistant Professor David J. Baker
Professor Laura Ruetsche
Professor Lawrence Sklar
Professor James D. Wells

For Lucy

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the members of my dissertation committee for numerous helpful conversations and comments on drafts of the chapters of this dissertation that greatly improved this dissertation. In particular, Gordon Belot has been amazingly generous with his time and patient as I worked and reworked many early drafts of this dissertation. I am indebted to him for pushing me to always produce better work and giving me the space to do so. David Baker, Laura Ruetsche, and Larry Sklar have been involved in this project from early on, and their comments and conversations have been invaluable. Dave, Gordon, Laura, and Larry also served as fantastic exemplars of how philosophy of physics ought to be done, and their careful and attentive work in the field has been an inspiration over the years. Finally, I am grateful for James Wells for his physicist's insight into the nature of symmetries.

I also owe a great debt to my fellow Michigan philosophy graduate students, most notably Dmitri Gallow, Lina Jansson, and Chip Sebens for their helpful discussion and comments on earlier drafts of many of the chapters of this dissertation. Thanks also to Jude Beck, Maureen Lopez, Lori Scott, and especially Linda Shultes for guiding me through the institutional bureaucracy of Michigan, and thanks to Molly Mahony both for her help in researching sections of this dissertation as well as her constant support. I am also especially grateful to those who put together a nice LaTeX template for me to use for my dissertation and to José Manuel Díaz Moreno, whose excellent YaLie package helped me through some of the technical difficulties in my project. Finally, I want to thank my friends, family, and most importantly, my wife, Lucy McNamara,

for supporting me through this work. I couldn't have done it without them.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Prospects for a New Account of Time Reversal

by

Daniel Peterson

Chair: Gordon Belot

Recent literature concerning the symmetry of time reversal has left unclear both what this symmetry is and what physical theories are invariant under it. In this dissertation, I argue that to understand how time reversal transforms physical states, we should seek symmetry transformations that meet minimal criteria of time reversal-hood under which all of the fundamental physical laws are invariant. If there is a unique transformation that emerges from this procedure, that transformation may be properly called time reversal. I apply my methodology to model worlds governed by various differential equations and examine the consequences of my account in each world. In the final chapter, I discuss the implications of failures of time reversal invariance in the fundamental laws for the nature of spacetime and conclude that two *prima facie* promising arguments for a connection between temporally asymmetric features of the laws and spacetime rest on questionable assumptions.

CHAPTER I

Introduction

Discussions of contemporary physical theories mention numerous physical symmetries, such as spatial translations, spatial rotations, and Lorentz boosts. But the focus of this dissertation is primarily on the time reversal symmetry transformation. What makes this symmetry so special that it should be the focus of this entire work, and, more generally, why exactly should we care about physical symmetries in the first place? My answer to this latter question may be found in the next chapter, but my answer to the first question unites the numerous chapters in this dissertation and explains how they fit together. Time reversal is a particularly important symmetry to philosophers who work on the metaphysics of time because our understanding of time is generally so poor. Contemporary philosophers of time still argue about basic, fundamental issues like the duration of the present and the existence of the past and future, suggesting that, despite years of inquiry, it is still difficult to make even incremental progress in understanding the nature of time. Physics provides us with a new perspective on some of these issues, and it seems likely that, by examining temporal symmetries like time translations and time reversal and their role in fundamental physical theories, we may gain a better understanding of what time is and what necessary and contingent features it has.

The unifying focus in this work is on the tantalizing suggestion that the failure of

the fundamental physical laws to be time reversal invariant gives us good reason to believe that spacetime itself may be temporally asymmetric which, in turn, may help explain the numerous and varied temporal asymmetries found in physical phenomena from entropy to causation to human experience and memory.[1] This suggestion has most recently, most clearly, and most forcefully been advocated by *Maudlin* (2007), but it has a long history. To my knowledge, it was first brought to the attention of contemporary analytic philosophers in the guise of the "Time Direction Heresy" discussed by *Earman* (1974), and something like this suggestion has appeared in works by *Horwich* (1987) and *Zeh* (2001), among others. This suggestion gives us reason to hope that a solution to one of the most difficult problems in the philosophy of time may be at hand, but I will argue that the hope it gives us is merely false hope. To thoroughly understand both the motivations for Maudlin's argument in particular and explore its significance, I assess two separate claims in this work: 1) that the fundamental laws of physics fail to be time reversal invariant, and 2) that the failure of the fundamental laws to be time reversal invariant would give us sufficient reason to posit a fundamental temporal asymmetry in spacetime.

The majority of this dissertation concerns an assessment of the first claim. In chapter 2, I pull back from the specific topic of time reversal invariance to give an account of physical symmetries more generally. I argue that, if we are to concern ourselves with physical symmetries that are useful for doing both the work of physics and the work of metaphysics, we ought to only consider symmetries that act consistently across physical theories. This consistency criterion provides the primary motivation for the account of time reversal I examine in chapter 3, where, after discussing and

---

[1]Throughout the course of this dissertation, I will treat the following claims as equivalent to one another: 1) Spacetime is temporally anisotropic; 2) Spacetime is temporally "handed" (see *Callender* (2000)); 3) Spacetime is temporally asymmetric. While there are subtle differences between the claims, they are not differences that matter for my analysis, and my argument that, for example, the failure of the fundamental laws of physics to be time reversal invariant does not support the temporal anisotropy of spacetime will support conclusions about spacetime's temporal asymmetry and temporal "handedness" as well.

rejecting what I call the "intuitive" and "theory-relative" accounts of time reversal, I argue for a promising new account of time reversal that avoids the problems of its competitors by steering a middle course between them. The mathematical underpinnings of this account are briefly discussed in chapter 4, and in chapter 5, I provide some toy models as test cases of my account and compare them to the verdicts of other accounts. In these chapters I argue that, in order to determine whether the fundamental laws of physics are time reversal invariant, an account of time reversal more mathematically involved than the ones currently available is needed, and I argue that, given my analysis of a few toy models, we have a good (if defeasible) reason to believe that the fundamental laws of physics fail to be time reversal invariant.

My final chapter of the dissertation thus returns to Maudlin's argument in particular and asks what to make of the previous chapters' results. My focus in this final chapter is primarily on the arguments Maudlin could rely on to justify his claim that, if the laws of physics fail to be time reversal invariant, this gives us a good reason to believe that there is some preferred temporal arrow in spacetime. Maudlin's claim here is not supported by further argument in his own work, so I draw on the work of *Earman* (1989) and *Brown* (2005) to provide two separate arguments for Maudlin's claim that seem, *prima facie*, fairly appealing. However, neither argument ultimately supports Maudlin's position, and so I argue that the consequences of physical laws' invariance (or lack thereof) under the time reversal symmetry may not be as metaphysically informative as we might have hoped. Even if the hope Maudlin gives us for explaining various "arrows of time" is false hope, the examination of why Maudlin's argument fails helps us to better understand physical symmetries generally, time reversal symmetries in particular, and the explanatory criteria that help us determine when we ought to adopt or accept a particular interpretation of a given physical theory, and these lessons will be important for other projects in philosophy of physics, philosophy of science, metaphysics, and even physics itself.

CHAPTER II

Physical Symmetries, Overarching Symmetries, and Consistency

Symmetries of various sorts play an extremely important role in our understanding of our best physical theories. Spacetime symmetries, dynamical symmetries, and empirical symmetries all tell us important information both about the world we live in and the physical theories we posit to explain and aptly represent that world. The focus of my project in this dissertation is on the symmetry of time reversal, one of the most important symmetries in physics. Physicists like *Carroll* (2010) frequently claim that the fundamental laws of physics are invariant under the symmetry of time reversal while at the same time acknowledging apparent violations of this symmetry at the fundamental level, such as the violation of CP invariance (and thus, by the CPT theorem, the violation of T invariance) by K-mesons. My purpose in this second chapter of my dissertation is to provide and justify a new methodology for determining whether a physical theory is time reversal invariant that will help to assess claims like Carroll's, but before discussing time reversal in particular, I will provide, in this chapter, a general account of two kinds of symmetries appealed to by physicists and justify a particular approach to investigating these symmetries that I will utilize later in the dissertation.

The word "symmetry" is bandied about in numerous contexts to various ends, so

it seems reasonable to begin a discussion of a symmetry like time reversal by isolating exactly what it is we generally mean by a "symmetry" within the context of the physical sciences. The following are examples of sentences we are likely to encounter involving the terms "symmetry" or "symmetric":

(1) Starfish have radial symmetry.

(2) Humans have bilateral symmetry.

(3) The "is equal to" relation is symmetric.

(4) The relationship between friends is more symmetric than the one
between parent and child.

(5) Time reversal is a symmetry of the fundamental laws of physics.

The focus of my discussion in this chapter will be on what symmetry means as it appears in sentences like (5) above, not as it appears in sentences like (1) to (4) (though, as we'll see, sentences like (1) to (4) may help us understand some general features of what I call physical symmetries); that is, I am interested in the notion of symmetry as it appears in our best fundamental physical theories. I divide such symmetries into two classes, namely a more fine-grained class of "physical" symmetries that are defined within the context of a specific physical theory, and the more coarse-grained "overarching" symmetries, which hold across numerous physical theories. My first task in this chapter is to lay out an account of physical symmetries, for while authors like *Belot* (2013), *Earman* (1989), and *Greaves* (2010) have provided their own formal treatments of similar symmetries, I hope to provide an account that better captures what it is for some transformation to be a symmetry *of* a particular physical theory.[1]

However, my focus in this work is on time reversal, an overarching symmetry, and as such most of this chapter will focus on overarching symmetries and the relation-

---

[1]My approach will ultimately be similar to Belot's but with a few important tweaks.

ship between overarching symmetries and the physical symmetries they connect. In particular, I will discuss "realism" and "conventionalism", two different stances one may take towards certain overarching symmetries, and provide several examples of how overarching symmetries, depending on which stance we take, may guide scientific inquiry. My central claim in this section is that should we wish to use overarching symmetries as guides to which new theories to accept or to the underlying metaphysical structure of the world, we ought to care most or only about overarching symmetries that act consistently across physical theories. The alternative stance towards overarching symmetries, which broadens the physicist's and philosopher of physics' domain of interest to include symmetries whose action on the same physical quantities may vary from theory to theory, does not allow overarching symmetries to do the work that both physicists and philosophers frequently require of them. I thus argue that both realists and conventionalists have good reason to prefer an account of time reversal that satisfies the realist's constraint of acting consistently across different physical theories.

2.1   What are Physical Symmetries?

Symmetries, generally speaking, are of interest because they reveal meaningful operations under which certain basic structures are preserved. In example (1) above, the starfish's radial symmetry tells us that, should the starfish be rotated by certain intervals about an axis passing through its stomach perpendicular to its legs, the starfish's shape will be preserved (or approximately preserved). The operation of interest in this example is a rotation, and the structure preserved is the starfish's shape. Likewise, in the case of bilateral symmetry, the figure's general shape is the relevant structure preserved by some transformation, but the operation of interest is different: bilateral symmetry picks out a structure that remains invariant under spatial reflection, unlike radial symmetry's reference to spatial rotations. And so in

fundamental physics as well, what counts as a symmetry is a kind of operation which leaves a certain structure invariant.[2]

As these examples suggest, there are essentially three features one must identify in order to understand exactly what a particular kind of symmetry is supposed to be: the relevant structure(s) left invariant by the symmetry, the objects acted on by the symmetry, and the way in which the symmetry transforms those objects. These three features of a symmetry are invariably linked. Consider spacetime symmetries, for instance. Spacetime symmetries are maps from spacetime points to spacetime points that are required to leave the spacetime interval (defined as $ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2$ for Minkowski spacetime) invariant. Dynamical symmetries are transformations that take a particular solution in the solution space of some set of equations to some other solution in that same solution space. Both spacetime and dynamical symmetries are thus maps from objects defined on a particular topological space (spacetime and the space of solutions respectively) to objects on that same particular topological space that leave some particular feature of those objects (the spacetime interval and solution structure respectively) invariant. Given that the first part of my project is to determine what sorts of transformations constitute physical symmetries, I should likewise explain what sorts of topological spaces and invariant structures are utilized by all physical theories.

2.1.1   A Formal Account of Physical Theories

Physical theories, as I use the term for the purposes of this paper, are essentially ordered tuples of the form $\langle E, X, U, N \rangle$, where E is the set of equations utilized by a particular theory (usually differential equations), X is the set of all independent variables appearing in the equations in E, U is the set of all dependent variables

---

[2]Something like this general notion of a symmetry can be found in numerous contemporary discussions of symmetries in the philosophy of physics literature. See, for instance, *Belot* (2013), *Brading and Castellani* (2007), *Healey* (2009), *Ismael and van Fraassen* (2003), *Roberts* (2008), and *van Fraassen* (1989).

appearing in the equations in E, and N is the "interpretation function" of that physical theory (more on that in a moment).[34] By dependent variables here I mean variables that can be represented as functions of the independent variables and whose derivatives in terms of these independent variables we take to be of physical interest. For instance, in Newtonian mechanics the variable representing the position of a ball rolling down an inclined plane can be represented as a function of the time that has passed since the ball was released, and so the variable "position" here is taken to be a dependent variable and "time" is taken to be an independent variable. There may be some conventional element in selecting which variables are dependent and which are independent for any particular physical theory; for instance, in the case of Newtonian mechanics, because the position x can be given in terms of the time t by $x = f(t)$, it is also the case that, for invertible f, $t = f^{-1}(x)$, so we may appeal to some conventional element to tell us whether time or position is the dependent variable here. Alternatively, we may take the velocity of the ball, $v = \frac{dx}{dt}$, to be more fundamental than its inverse $\frac{dt}{dx}$ and so take there to be some principled, non-conventional reason for taking t as the independent variable here instead of x. Typically, the variable which is easier for experimenters to control or manipulate is taken to be the independent variable, but this need not always be the case. My point here is only that it is up to the theory we are interested in to tell us what quantities we care about and how to assign dependent and independent variables in any given situation.

The equations $e \in E$ on my account are conditions on functions of independent variables, dependent variables, and, in the case of differential equations, derivatives of

---

[3]I should note that I am not trying to give a full account of what it is for something to count as a physical theory; rather, what I refer to here as a "physical theory" is a set of necessary features that I take all physical theories to have and all of which I will exploit in my account of symmetries later in this paper.

[4]Over the course of this paper I will refer to X as both the set of independent variables and the space whose axes correspond to the independent variables in X. The context will help to determine which of the two I mean at any given time, and I will similarly equivocate with my usage of U.

the dependent variables with respect to the independent variables that require these functions be zero.[5] So, for instance, in cases where there is only one independent variable x and one dependent variable u, a differential equation can be represented as the condition that the function $d(x, u, u^{(1)}, ..., u^{(n)}) = 0$, where $u^{(i)}$ is the nth derivative of u with respect to x (e.g. $u^{(2)} = \frac{\partial^2 u}{\partial x^2}$).[6] An equation can be characterized, then, by the function it sets equal to zero, and so I will take the elements of the set E to be the functions that the equations of the physical theory set equal to zero. Differential equations serve as constraints on the dependent and independent variables utilized by a theory and help us determine how the dependent variables depend on the independent variables of a theory, but they are more difficult to deal with than algebraic equations which do not depend on any of the derivatives of the dependent variables. Unfortunately, physics is complicated, and so many of our most successful physical theories postulate differential equations instead of algebraic equations and are better at representing, predicting, and explaining the world for it.

Differential equations and the variables they utilize would fail to interest physicists if these mathematical objects did not tell us anything about the physical world; however, as it turns out, many important physical systems are aptly modeled by differential equations, and, as previously stated, the differential equations we appeal to in our physical theories have been extremely good at predicting and explaining various physical phenomena. Differential equations allow us to represent physical constraints because we take the variables appealed to by our physical theories to represent measurable features of the physical world. Physical theories necessarily contain what I will call an interpretation function N, which is a map from the variables appealed to by a physical theory and functions of these variables to the set of measurement processes in the physical world that provide the values for these variables. Interpreta-

---

[5]This account follows the one given in *Olver* (1993).

[6]Throughout this paper I will use D instead of E and d_i instead of e_i when speaking of a set of differential equations and a particular differential equation respectively.

tions are necessary components of physical theories because two physical theories may utilize similar equations and variables to different ends. For instance, the differential equation:

$$d(x,t,u) = \frac{\partial^2 u}{\partial^2 t} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \qquad (2.1)$$

is the one-dimensional wave equation and can be used to model different kinds of waves, including compression waves (like sound) and transverse waves (like light). The context in which we find this differential equation, then, tells us a great deal about its content, and we need that information if we want to use this differential equation to make useful predictions. The interpretation function helps us in this case by telling us both which measurements will give us the values we need to implement the wave equation and how these variables are similar to (or differ from) variables that show up in other physical theories. The function N can be as complicated or simple as one likes, but at the very least it must provide a link between variables and the physical measurements we perform to determine the values of those variables when using a theory's equations. I want to emphasize here that I'm not providing a theory of where N comes from, only stating what it does (tells us which variables represent which quantities) and assuming that, for any given physical theory, we have some N readily available to us for our purposes. The interesting question of just how one determines what N is in the context of a particular physical theory lies beyond the scope of this dissertation.

One may object here that physical theories do not all come equipped with particular interpretations; after all, many physical theories (most notably quantum mechanics) have numerous interpretational difficulties, and so stipulating that a physical theory like quantum mechanics comes with "an interpretation" seems strange. Note, however, that my technical usage of the "interpretation function" differs from what is typically meant by an interpretation of a physical theory in other contexts: the job of

my "interpretations" is only to tell us what measurement processes give us the values for the variables postulated by a physical theory, not to tell us anything about the underlying metaphysics of the world described by this theory. Note also that I have not stipulated anything about what sort of map N must be. This is simply because not all of the variables utilized in our equations correspond directly to measurable features of the world. The quantum mechanical wave function, for instance, may not be directly measurable, but the Born rule tells us that its modulus squared corresponds to the probability of finding the system described by the wave function in a certain state. Assuming that probabilities are empirical features of the world (which itself is something of a contentious assumption), it is clear that, even if the theory of quantum mechanics has no agreed-upon interpretation in the traditional sense of the word, it does have an agreed-upon N in my technical sense.

Let's turn from my discussion of what physical theories are to the question of what topological space the central objects of the physical theory live in. For starters, I will examine the simple case of a physical theory all of whose equations are algebraic equations; that is, I will restrict my consideration only to physical theories containing sets E such that:[7]

$$\forall e_i \in E, \exists f \text{ such that } e_i = f(X, U) \tag{2.2}$$

That is, assume that the equations in question depend only on the dependent and independent variables, not their derivatives. Such equations are represented by curves on a given manifold, namely the space of dependent and independent variables X and U which I will henceforth call the *variable space* $X \times U$.[8] Each point in this space corresponds to a set of values for the variables in X and U, and the theory

---

[7]To be precise, we require here that f be a differentiable function.

[8]I should mention that, while I call the representations of these equations "curves", they are more correctly n-1-dimensional submanifolds of n-dimensional manifolds. For instance, in cases where the variable space is greater than 2-dimensional, the "curve" of the equation will likewise be greater than 1-dimensional, standard terminological implications to the contrary.

differentiates between those sets of values that are *physically possible* according to a particular equation, which correspond to points on the equation's curve, from those sets of values which are not physically possible according to that equation. Few theories contain singleton sets of equations, and so typically we are interested not in each curve on its own but rather the intersections of the curves corresponding to all of the equations contained in a particular physical theory. It is these curves that determine which sets of variable values the theory takes to be physically possible simpliciter and which values it does not. Relative to this curve, then, we can define a *solution* of an algebraic equation as a map from the space of independent variables X to the space of dependent variables U such that, when this map is identified in the obvious way with a subset of the variable space $X \times U$, this subspace lies entirely within the curve determined by the equations. This notion of a solution extends naturally to characterize solutions of differential equations as well, as we'll see in the next section.

2.1.2 Physical Symmetries and How We Find Them

Following this definition of solution, one can now posit a notion of symmetry that lines up nicely with the previously discussed notion of a dynamical symmetry. A symmetry of a physical theory T is a transformation that maps points in T's variable space to other points in T's variable space that keep the solution structure the same.[9] That is, a symmetry cannot map the points in a subspace that lies entirely within the curve determined by T's equations to points outside of the curve determined by T's equations. The objects transformed by a physical symmetry are thus points in a theory's variable space, and the structure preserved by a physical symmetry

---

[9]I will restrict my talk of symmetries in this paper to point symmetries. Generalized physical symmetries may be of interest as well, and my account of point physical symmetries may extend to an account of generalized physical symmetries, but as time reversal is a point symmetry, I will restrict my notion of a physical symmetries to only point symmetries for the purposes of this paper. Those interested in generalized symmetries should consult chapter 5 of *Olver* (1993).

transformation is the solution structure of that space.

The set of all transformations of a physical theory, along with a binary operation that defines the product of any two such transformation, composes a *group*, meaning that a symmetry group G_T can be defined as the set of all transformations of a physical theory T with a binary operation $*$ (in this case, functional composition) satisfying the following four conditions:

1. Closure: $\forall A, B \in G_T,\ A * B \in G_T$
2. Associativity: $\forall A, B, C \in G_T,\ A * (B * C) = (A * B) * C$
3. Identity: $\exists e \in G_T$ such that $\forall A \in G_T,\ e * A = A * e = A$
4. Inverse: $\forall A \in G_T,\ \exists A^{-1} \in G_T$ such that $A^{-1} * A = A * A^{-1} = e$.

In essence, each of these conditions says the following: from 1, the product of any two symmetry transformations must itself be a symmetry transformation; from 2, the binary operation utilized by the group is associative; from 3, the identity operation (i.e. the operation that maps every point in a theory's variable space to itself) must always be a member of a group of physical symmetries; and from 4, every symmetry transformation has an inverse that is also a symmetry transformation.

The search for physical symmetries is thus just the search for the group of symmetries that leaves the solution structure of a variable space invariant. I will forego some technical details here, but there are available mathematical methods that take advantage of the infinitesimal generators of groups that allow us to determine what the symmetry group of any particular differential or algebraic equation is.[10] So, in short, once we restrict our attention to physical theories that utilize only algebraic equations and define physical symmetries as I have, there are mathematical results that make the calculation of these symmetries (relatively) easy in many cases. So far,

---

[10]Those who would like more detail on just how this process works for the case of symmetries continuously connected to the identity transformation should consult Section 2.1 of *Olver* (1993) and chapter 4 of this dissertation.

so good.

The problem is that, as stated previously, most physical theories consist of differential, not algebraic, equations, and unlike algebraic equations, differential equations cannot be represented as curves on a theory's variable space. So how do we extend the notion of a symmetry group, which was so useful in the case of algebraic equations, to the case of differential equations? The answer is that, as in the algebraic case, we must find some geometric object (like a curve) defined on a manifold (like variable space) that corresponds to the vanishing of some function. Differential equations, like algebraic equations, can be represented by vanishing functions, albeit of a slightly different sort. A differential equation is just an equation

$$d(X, U, U^{(1)}, ...U^{(n)}) = 0 \tag{2.3}$$

where $U^{(i)}$ is the set of all ith-derivatives of each variable in U with respect to each variable (and each combination of variables) in X. So, for instance, if a theory has two independent variables $x, y \in X$ and one dependent variable $u \in U$, $U^{(2)}$ is given by the following:

$$U^{(2)} = \left\{ \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial x \partial y}, \frac{\partial^2 u}{\partial y^2} \right\}. \tag{2.4}$$

If we're willing to treat each of these derivatives as its own variable (which we can do so long as we impose some constraints later on in our process), we can take each differential equation to represent a curve in the space $X \times U \times \cdots \times U^{(n)}$ for some fixed n. Such a space is said to be the *nth prolongation* of the base space $X \times U$ since this variable space can be identified with a subspace of the prolongation space. The point here is just that, as required, we can now represent differential equations as curves on a manifold (the nth prolongation of variable space), which means that the previously discussed symmetry group methods are applicable again.

One final result, which I will only state but not prove, is that, for each symmetry group of the nth prolongation of the variable space G_p, there is a corresponding symmetry group of the variable space G_v such that G_p is the prolongation of G_v.[11] What this means, in essence, is that every element of G_v is a transformation of variable space that induces a transformation of the nth prolongation of that variable space in G_p and such an element of G_v exists for every element of G_p. So, once we know the symmetry group of a differential equation (or set of differential equations) in the prolongation of variable space, we can work backwards to find a group of transformations of the original variables in X and U such that each transformation of X and U in this group takes solutions of that differential equation (or set of differential equations) to other solutions of that differential equation (or set of differential equations).[12]

If my account of physical theories and physical symmetries is correct, then we have an intuitively simple (if complicated to implement) procedure for determining all of the physical symmetries of a given physical theory: determine the equations and variables of interest to the theory you want to consider, and then use the method outlined above to determine all of the physical symmetries of those equations. In the next section, I'll discuss how this analysis helps us determine the form of "overarching" symmetries like time reversal, which in turn will suggest my methodology for the rest of this project.

---

[11]Again, parties interested in the technical details are referred to chapter 2 of *Olver* (1993).

[12]I should note here that, as sketched above, this method only gives us all of the symmetries of a physical theory that are continuously connected to the identity transformation, which may leave some scratching their heads since my investigation concerns a *discrete* symmetry, namely time reversal. In a series of publications, *Hydon* (1998a,b, 2000a,b) provides a method of determining all of the discrete symmetries of a differential equation from the symmetry group obtained using the methods I outlined above. I will use Hydon's method in later chapters to determine all of the discrete symmetries of a series of differential equations from these differential equations' continuous symmetry groups.

2.2   The Problem of Overarching Symmetries

My account of physical theories and symmetries in the last section gives us some idea of what it is to be a symmetry of a certain physical theory; however, there are plenty of contexts in which one may speak of a symmetry without referring to a particular physical theory. For instance, physicists may speak of translations, rotations, boosts, time reversal, parity reversal, and the consequences of these symmetries without making reference to any one theory in particular. What's more, symmetries can sometimes pull double-duty, both arising as a consequence of, say, the background spacetime of some theory and serving as a constraint on the dynamical equations of that same theory. The classical Klein-Gordon theory provides an example of symmetries pulling such double duty since the spacetime symmetries that arise from taking Minkowski spacetime as the background for the theory serve as constraints on the classical Klein-Gordon Lagrangian and the Klein-Gordon equation itself.

Such examples lead to the following problem for any account like mine that utilizes both theory-relative and theory-independent symmetries, which I call the "Problem of Overarching Symmetries": on one hand, my account tells us that physical symmetries are defined relative to a particular theory since they are transformations defined on the theory's variable space or prolongation of the theory's variable space, but on the other hand, we seem quite capable of extending our notion of some particular physical symmetry beyond the theory invariant under it to identify two seemingly different transformations on different variable spaces as one and the same. What, then, is the relationship between theory-relative symmetries (physical symmetries) and theory-independent symmetries (overarching symmetries)? My statement of this problem is a bit abstract, so let's look at an example: classical Newtonian gravity and classical electromagnetism are two different physical theories with different sets of dependent and independent variables, and the two theories are governed by different sets of equations; however, despite these differences, both theories are said to be invariant

under a number of the same symmetries, such as spatial translations and rotations. So we have a set of symmetries called "spatial translations and rotations" that, in various contexts, may seem to pick out 1) point symmetries in Newtonian gravity's variable space and 2) point symmetries on classical electromagnetism's variable space. We lump these two "fine-grained" physical symmetries under a single heading and associate two different physical symmetries with one and the same "overarching symmetry", which is just a map that assigns a particular physical symmetry or a set of physical symmetries to every theory. But the question still remains: in what sense can we say that Newtonian gravity and classical electromagnetism are invariant under "the same" symmetries on my account? And what is the strength of this "sameness"? Though I will not answer this question in full, my analysis in the remainder of this chapter will shed some light on how we treat or ought to treat these overarching symmetries.

Overarching symmetries clearly need to be treated differently from physical symmetries, but before saying more, I should say a bit about the importance of overarching symmetries. Later on in this chapter I will give examples of the kind of work overarching symmetries can do for physicists and philosophers of physics, but despite their usefulness, I take overarching symmetries to be less fundamental than physical symmetries. Overarching symmetries can be thought of as ways of picking out properties of interest that relate symmetries we care about. We might claim that what unites some set of physical symmetries under the overarching symmetry of, say, spatial translation is that all of these symmetries map solutions of a theory to solutions that differ from the first solution only in their spatial coordinates in that theory's variable space, or we might say that the relevant property that unites spatial translations is that it performs this mapping without changing any other features of the mapped solution. I will leave aside for now the question of whether there is any *natural* property *á la Lewis* (1983) picked out by all overarching symmetries and

instead focus on the pragmatic question of which properties are the most practical for physicists and philosophers to use. My account thus allows that the question "Is X really an overarching symmetry?" may or may not be substantive; the real project for those investigating overarching symmetries (as opposed to physical symmetries), I contend, isn't to figure out what overarching symmetries correspond to the *real* overarching symmetries in the world (if there are any) but rather to lay down a useful criterion for what constitutes a particular overarching symmetry and argue why this criterion is the most useful one.

So, even if we treat physical symmetries rather than overarching symmetries as the more fundamental objects and admit that the question "which overarching symmetries are the *real* overarching symmetries?" may not be substantive, the following problem still remains: when should we count two physical symmetries as corresponding to the same overarching symmetry? Or, put more formally, the problem of overarching symmetries can be cast as follows: suppose that some theory $T_1 = \langle E_1, X_1, U_1, N_1 \rangle$ with n independent variables and m dependent variables is invariant under the point symmetry transformation S_1, which operates on a point $p = (x_1^1, \ldots, x_1^n, u_1^1, \ldots, u_1^m)$, where $x_1^i \in X_1$ and $u_1^i \in U_1$, as follows: $S_1(p) = (f_1(x_1^1, \ldots, x_1^n, u_1^1, \ldots, u_1^m), \ldots, f_{n+m}(x_1^1, \ldots, x_1^n, u_1^1, \ldots, u_1^m))$ for some functions $f_1, \ldots, f_{n+m}$. Now, suppose there is another theory $T_2 = \langle E_2, X_2, U_2, N_2 \rangle$ with k independent variables and l dependent variables and points $(x_2^1, \ldots, x_2^k, u_2^1, \ldots, u_2^l)$ which is invariant under the point symmetry transformation S_2, which operates on a point $q = (x_2^1, \ldots, x_2^k, u_2^1, \ldots, u_2^l)$, where $x_2^i \in X_2$ and $u_2^i \in U_2$, as follows: $S_2(q) = (h_1(x_2^1, \ldots, x_2^k, u_2^1, \ldots, u_2^l), \ldots, h_{k+l}(x_2^1, \ldots, x_2^k, u_2^1, \ldots, u_2^l))$ for some functions $h_1, \ldots, h_{k+l}$. Under what conditions can we say that T_1 and T_2 are invariant under the same overarching symmetry transformation?

The problem perhaps wouldn't be so serious if it weren't for the fact that variables that appear in two different physical theories may correspond to the same feature of

the physical world or may be calculated by exactly the same methods. We may refer to the length of a metal bar both in the context of the classical theory thermodynamics when we heat up one end of the bar and want to know how long it will take for the temperature to rise by a certain amount at the other end of the bar, and we may refer to it in the context of the theory of special relativity when we accelerate the metal bar to some near-light velocity and want to determine the effects of length contraction. In both cases the length of the bar can be calculated by the same sorts of measurements because the physical quantity, length, is the same in each case. In essence, we may be especially worried about how to pick out the symmetry corresponding to S_2 in the formalism above when, for some $a \in X_1 \cup U_1$ and $b \in X_2 \cup U_2$, $N_1(a) = N_2(b)$ or $N_1(f(a)) = N_2(f(b))$ for some f.

Though I cannot address all of the criteria one could lay down as good candidates for determining which physical symmetries fall under the same overarching symmetry, I can outline two positions that one could take regarding these overarching point symmetries based on the criterion that I will use for the purposes of my project. The difference between these two positions requires us to posit the following definition of what I will call "interpretive identity", which I take to be a necessary (but not always sufficient) condition for saying that two variables correspond to the same physical property:

> (*Interpretive Identity*): Take $T_1 = \langle E_1, X_1, U_1, N_1 \rangle$ and $T_2 = \langle E_2, X_2, U_2, N_2 \rangle$
> to be two different physical theories. Two variables $v_1^i \in X_1 \cup U_1$ and
> $v_2^j \in X_2 \cup U_2$ are said to be *interpretively identical* just in case $N_1(v_1^i) =$
> $N_2(v_2^j)$ or $N_1(f(v_1^i)) = N_2(f(v_2^j))$ for some function f.[13]

---

[13] A more complicated (and likely correct) theory would also utilize a property of *Approximate Interpretive Identity* (*AII*) in much the same way. We could define *AII* as follows: Take $T_1 = \langle E_1, X_1, U_1, N_1 \rangle$ and $T_2 = \langle E_2, X_2, U_2, N_2 \rangle$ to be two different physical theories. Two variables $v_1^i \in X_1 \cup U_1$ and $v_2^j \in X_2 \cup U_2$ are said to be *approximately interpretively identical* just in case $N_1(f_1(v_1^i)) = N_2(f_2(v_2^j))$ for some functions f_1 and f_2. The consistency definition would then carry the additional constraint that approximately interpretively identical variables are transformed in the same way; that is, for any two approximately interpretively identical variables x_1 in T_1 invariant

Essentially, two variables are said to be interpretively identical if those variables or the same function of those variables are mapped by each theory's interpretation function to the same measurement processes. With this definition under our belts, we can formulate the following definition:

> (*Consistency*): Take $T_1 = \langle E_1, X_1, U_1, N_1 \rangle$ to be a physical theory invariant under the symmetry transformation S_1, which maps each coordinate $v_1^i \in X_1 \cup U_1$ to $f_{v_1^i}(v_1^1, v_1^2, \dots)$, and take $T_2 = \langle E_2, X_2, U_2, N_2 \rangle$ to be a physical theory invariant under the symmetry transformation S_2, which maps each coordinate $v_2^j \in X_2 \cup U_2$ to $f_{v_2^j}(v_2^1, v_2^2, \dots)$. Take the two sets of variables $A \subseteq X_1 \cup U_1$ and $B \subseteq X_2 \cup U_2$ to be such that $\forall v_1^i \in X_1 \cup U_1$, if $\exists v_2^j \in X_2 \cup U_2$ such that v_1^i and v_2^j are interpretively identical, then $v_1^i \in A$, and $\forall v_2^j \in X_2 \cup U_2$, if $\exists v_1^i \in X_1 \cup U_1$ such that v_2^j and v_1^i are physically identical, then $v_2^j \in B$. S_1 and S_2 are *consistent* if and only if, $\forall a_i \in A$ and $\forall b_i \in B$, the interpretive identity of a_i and b_i implies $f_{a_i}(v_1^1, v_1^2, \dots)$ $= f_{b_i}(v_2^1, v_2^2, \dots)$ for some fixed values of the variables $v_1^i \notin A$ and $v_2^j \notin B$ for which f_{a_i} and f_{b_i} are not constant functions.

Despite its formal complexity, *Consistency* is intuitively easy to understand. Essentially, two physical symmetries are consistent only if both symmetries treat the "same variables" in the "same way". By "same variables" here I mean variables that are interpretively identical, and by treating these variables in the "same way", I mean that, ignoring any variables that aren't interpretively identical, S_1 treats the variables in $X_1 \cup U_1$ in the same way that S_2 treats those variables' interpretively identical counterparts. To give an example: suppose that our first theory has a symmetry S_1 that transforms its points as follows: $S_1(x, y, z) = (z(x + y), y, z)$, and suppose

---

under S_1 and x_2 in T_2 invariant under S_2, S_1 and S_2 would be consistent only if $f_1(S_1 x_1) = f_2(S_2 x_2)$. I have excluded this extra criterion from my main treatment because it extends quite naturally from the account I provide and I don't wish for an already complicated definition to be made more complicated than it needs to be.

that our second theory has a symmetry S_2 that transforms its points as follows: $S_2(x, y, w) = (x + y + w, y, w)$, where I have used the same variable name to denote variables in different theories that are interpretively identical with one another. S_1 and S_2, according to my definition, are consistent since, in cases where $z = 1$ and $w = 0$, $S_1(x, y, z) = (x + y, y, 1)$ and $S_2(x, y, w) = x + y, y, 0)$; however, S_3, which acts as follows: $S_3(x, y, w) = (x + y, x + y, w)$, is not consistent with S_1 per $Consistency$ since it is not generally the case that $y = x + y$.

$Consistency$ has the form it does for several reasons. First, it requires that the sets A and B consist of all variables in one theory with interpretively identical counterparts in the other theory. Otherwise, physical symmetries of two theories that treat the "same" variable x consistently but not the "same" variable y might mistakenly be called consistent with one another. What $Consistency$ tries to capture is not merely consistency relative to some subset of variables or another but rather consistency with respect to all variables interpretively identical to variables in the other theory under consideration. Secondly, it requires that consistent symmetries map interpretively identical variables to the same function of those interpretively identical variables modulo some variables that have no interpretively identical counterparts in the other theory considered. These variables with no interpretively identical counterparts in the other theory may sometimes make it hard to determine whether two different symmetries are acting consistently, so $Consistency$ requires that the identity of the outputs of the two functions f_{a_i} and f_{b_i} hold only for at least one case where the variables without interpretively identical counterparts take on constant values.[14] We

---

[14]This feature of $Consistency$ makes it fairly weak constraint, and we might prefer to accept a constraint that requires variables that appear in one theory but not another to assume some particular value when comparing the symmetries of the two theories, perhaps due to the fact that one theory is "approximately" equivalent to the other when this extra variable takes on a particular value. For instance, if we treat the speed of light c as a variable in special relativity, we may require that symmetries in Newtonian mechanics and special relativity can only be called "the same" symmetry if they transform the same variables the same way in the case where c is idealized as infinitely large. Such a constraint would be reasonable, but as a discussion of this extension of $Consistency$ would lead to a detailed, albeit interesting, digression, a discussion of this constraint will have to wait for another time. However, note that adding this constraint would still make one

can ignore cases where these variables not in A or B are added to otherwise identical functions by setting these variables to zero, and we can ignore cases where these variables not in A or B are multiplied by otherwise identical functions by setting these variables to one. Finally, *Consistency* requires that the functions not be constant functions to rule out cases where, for instance, one might claim that the symmetry S_1, which takes the variable x to wxy and S_2, which takes the variable x to $z(x+y)$ are consistent because, when w and z are zero, these two functions are the same (more specifically, both functions are equal to 0).

I find the above treatment of variables in one theory not interpretively identical to variables in another theory adequate for many of the cases I have in mind, but I admit that some may reasonably object to my treatment because of cases like the following: imagine that we have two theories, T_1 and T_2, the first of which is invariant only under the symmetry transformation $S_{1c}(x,t) = (x+c, t)$ and the second of which is invariant only under the symmetry transformation $S_{2c}(x, y, z, t) = (y^4 x + \frac{c}{z}, y, z, te^{1-y} + z - 1)$, where x and t are interpretively identical across the two theories but y and z have no counterparts in T_2. S_{1c} and S_{2c} are consistent in my sense of the term because $S_{2c}(x, 1, 1, t) = (x+c, 1, 1, t)$. But, my objectors say, this is crazy! Look at the strange y- and z-dependent changes that S_{2c} induces. Given such strange changes, how can we call S_{1c} and S_{2c} consistent symmetries in any sense?

My response to this objection is, in part, to bite the bullet: if one has the intuition that S_{1c} and S_{2c} cannot be consistent and that any theory of overarching symmetries that treats them as such is wrong, theres little I can do to make my account seem compelling. I will say that, absent any concrete physical theories to examine in this toy example, we may be ignoring features that may make the consistency between the two symmetries more palatable in a more fully fleshed-out case. For instance, it may

---

a realist in my sense of the term, meaning that nothing in my arguments for the rest of this chapter relies on the inclusion or lack of this constraint in *Consistency*. I will have more to say about the case of velocity boosts in special relativity and Newtonian mechanics shortly.

be the case that T_2 is the more fundamental theory than T_1 and the new variables y and z tell us something important about the structure of our variable x that leads us to think of S_{1c} as an idealized version of S_{2c}. Absent more details about T_1 and T_2, this is, of course, mere posturing, but it doesnt seem unreasonable to me to think that there are many contexts in which calling S_{1c} and S_{2c} consistent would be perfectly reasonable.

But lets assume that the objector has a more principled objection than just that calling these symmetries consistent seems weird. Perhaps the objector thinks that the deal-breaker here is the fact that the coordinate t is unchanged by S_{1c} in T_1 but is changed by S_{2c} in T_2, albeit by variables with no interpretively identical counterparts in T_1. My consistency criterion, the objector thinks, should be amended to require that two consistent symmetries transform the same variables and leave the same variables unaffected. This seems a *prima facie* well-motivated amendment, but I take the following case to give us a reason to leave my consistency criterion as-is.

Consider Newtonian mechanics and special relativity. Newtonian theories are invariant under Galilean boosts, which act on the Newtonian variable subspace (x, t, u) as follows: $G_v(x, t) = (x + vt, t, u + v)$, where x is position, t is time, and u is velocity. The Lorentz boosts under which special relativistic theories are invariant act on the variable subspace (x, t, u) as follows: $L_v(x, t, u) = ((x - vt)/\sqrt{1 - \frac{v^2}{c^2}}, (t - \frac{vx}{c^2})/\sqrt{1 - \frac{v^2}{c^2}}, \frac{v+u}{1+(\frac{vu}{c^2})})$, where c is the speed of light. Note that, while G_v leaves t unchanged, L_v indeed changes t. As written, these two symmetries are quite different from one another and are inconsistent. However, if one adopts a broad reading of my consistency criterion and takes $\frac{1}{c}$ to be a "variable" in special relativity with no interpretively identical counterpart in Newtonian mechanics, look at the result on the parameter (not variable, since c is a constant) subspace $(x, t, u, \frac{1}{c})$: $L_v(x, t, u, 0) = (x + vt, t, u + v, 0)$. That is, as any good first year physics undergraduate knows, when we idealize c away by setting $\frac{1}{c} = 0$, L_v and G_v transform

the interpretively identical variables consistently. There seems to be a sense in which Galilean boosts and Lorentz boosts *are* actually consistent. But again, this is obvious: theyre both velocity boosts! Galilean boosts tell you how to transform the coordinates of some event in your inertial frame so that theyll agree with the coordinates assigned to that same event by a traveler in a different inertial frame in the context of Newtonian mechanics, and Lorentz boosts tell you exactly the same thing in special relativity. So I take it to be a welcome consequence of my account that, depending on what we think of the role of c in special relativity, we may conclude that Lorentz boosts and Galilean boosts bear important similarities to one another despite looking so different from one another. But note that this conclusion could not be drawn if we were to amend *Consistency* in the way my objector suggests since Lorentz boosts dont leave t unaffected as Galilean boosts do. It seems to me, then, that despite having to accept some strange consequences like the consistency of S_{1c} and S_{2c}, we have good reason to accept the account of *Consistency* I have proposed unamended.

At least two positions can be defined relative to my notion of consistency: those who require that all physical symmetries associated with some overarching symmetry be consistent or who adopt some equally strong or stronger criterion of overarching symmetry-hood I will call "realists", and those who deny consistency as a constraint on uniting physical symmetries under the label of some overarching symmetry and adopt some weaker criterion I will call "conventionalists"[15]. Realists hold, basically,

---

[15]Note that, as defined, realism and conventionalism are relative to the particular overarching symmetry under consideration. I see no reason to disallow, say, realism about time reversal and conventionalism about gauge transformations, supposing one has good reason for treating these symmetries differently.

Also, for the purposes of this paper, I will take *Consistency* to be realists' necessary criterion of overarching symmetry-hood, though as I state, any equally strong or stronger criterion would qualify as a realist position as well. I believe there are good reasons to take consistency to be the defining criterion of overarching symmetries even over other realist symmetries (for instance, the realist criterion that requires that physical symmetries be consistent *and* act on the same variable space, which seems to limit the notion of an overarching symmetry to the point where it is no longer useful or applies to many motivating cases for overarching symmetries like classical mechanics and classical electrodynamics as discussed above), but my target in this work is the conventionalist, or

that some overarching symmetry (or, in its most radical form, all overarching symmetries) must be a transformation that acts consistently (or at least in accordance with some criterion at least as strong as *Consistency*) across physical theories in all cases. *Consistency* provides a necessary (though not necessarily sufficient) condition for uniting two physical symmetries under the same overarching symmetry. For instance, if two different theories both refer to, say, time[16], then if one theory is invariant under the physical symmetry "time translation" that transforms the time coordinate t by taking it to the coordinate $t + a$, then the second theory can only be invariant under the same overarching symmetry of "time translation" if it is invariant under a transformation that likewise takes t to $t + a$ (modulo some variables that may appear in one theory but not the other). More importantly, if the two theories refer to some common set of parameters (e.g. time, position, and momentum), then the two theories are invariant under the same overarching symmetry, according to the realist, only if there is some symmetry transformation defined on the first theory that treats time, position, and momentum the same way that some symmetry transformation defined on the second theory treats these parameters.

Conventionalists, on the other hand, reject consistency or any equally strong or stronger constraint as a constraint on overarching symmetries and embrace something weaker, making it easier for two different symmetries to be identified with the same overarching symmetry. For instance, in the previous example, some sort of conventionalist could argue that, despite the fact that, say, one theory is only invariant under a symmetry that takes t to $t + a$ and the other theory is only invariant under a symmetry that takes t to $-t$, we may still consider these two different theories to be

---

at least the conventionalist who doesn't believe in investigating consistent symmetries first. I leave it to other realists who believe my definition of *Consistency* to be too weak to provide and defend an alternate criterion for overarching symmetries.

[16]I will not speculate in this paper on just how to determine whether one of the variables appealed to within the context of a particular physical theory is, in fact, time, but two promising proposals for what makes time distinctive which are similar in spirit (if not in the stances they take regarding lawhood) are put forward by *Skow* (2007) and *Callender* (2011).

related by a single overarching symmetry because both symmetries satisfy the weaker criterion of changing the same variable. Conventionalists can adopt any number of constraints, some of which may be quite close to consistency. For instance, the conventionalist may adopt a criterion that physical symmetries must act consistently on one particular variable only in order to be united under one overarching symmetry, or they may require that all physical symmetries that act in a particular way on one particular variable may be united under the same overarching symmetry. Regardless of the specific criterion, however, what separates realists and conventionalists is that conventionalists allow strictly more physical symmetries to qualify as potential instances of a particular overarching symmetry than realists allow.

The difference between realism and conventionalism becomes salient when trying to determine certain important features of physical theories. We may be interested, as many philosophers of time and philosophers of physics have been, in the question of whether or not the fundamental laws of physics are invariant under the time reversal operator T. Realists trying to answer this question may proceed quite differently from conventionalists. Realists will look at each fundamental physical theory in question to determine whether or not its time reversal symmetry (if it exists) can be unified with the symmetries of the other fundamental physical theories consistently. Conventionalists, meanwhile, will likewise examine each physical theory to be sure that there is some point symmetry suitably called "time reversal" (under some weaker standard) under which it is invariant, and should they find such a transformation (or such transformations) for all fundamental physical theories, they will be happy to agree that the fundamental laws of physics are invariant under time reversal. For the rest of this chapter, I will refer to the general strategy of looking for an overarching symmetry that acts consistently across a number of different physical theories as the realist strategy and the strategy of broadening this search for an overarching symmetry's physical manifestation within a particular theory to include operations

that differ significantly from theory to theory as the conventionalist strategy. I should point out that conventionalists can adopt the realist strategy just as easily as realists can and may be led to do so if we have reason to think that consistent convention-alist overarching symmetries have some advantage over inconsistent conventionalist overarching symmetries. It is to the realist strategy that I refer when I say that we ought to "act like realists" even if we are not.

Realism is a stronger and more severe stance towards the identity of symmetry transformations across theories than conventionalism is, and as such we can expect that realism will make the invariance of a set of theories under some particular sym-metry harder to obtain than conventionalism would make it. So, when we ask the question of whether or not the fundamental laws of physics are time reversal invari-ant, for instance, which strategy should we adopt, or perhaps more importantly, which one should we adopt first? The answer seems fairly clear (to me, at least): adopt the stronger standard at first (that is, use the realist strategy), and should we find that there is no single symmetry under which all of the laws in question are invariant, perhaps then we can fall back to the weaker conventionalist stance and try to find some notion of time reversal invariance there.

My intuition in this case follows from the more general philosophical strategy that, when faced with an easier question and a harder version of the same question, tackling the harder version first will usually be more enlightening than tackling the easier question first would be. There are many other cases where our intuitions tell us that we ought to consider one particular theory instead of its rival(s). For instance, though this may be a minority opinion, I agree with *Belot* (2011, Appendix A) that the simplicity of a physical theory may not be a mark in favor of that theory's truth, but it gives us a good reason to test this theory first. After all, simpler theories are typically easier to test, and so adopting simpler theories first will allow us to test a greater number of theories in the same amount of time than we would have if we tested

only more complicated theories. We are typically better off testing more unifying and universal theories first as well. If we find three white swans, we are inclined to test the hypothesis "All swans are white" before we weaken this hypothesis to "Most swans are white", which is harder to disconfirm.

I do not mean to rest the bulk of my argument in favor of acting like a realist on this particular intuition that theories with certain theoretical virtues are the ones we ought to adopt and/or test first, but I do believe that this intuition may provide a strong point in favor of realism among those who have intuitions similar to mine. Still, it remains to be shown that the realist strategy, when applied to time reversal invariance, for instance, is likely to yield better and more useful results than the conventionalist strategy. In the remainder of this chapter, I will argue that the realist strategy ought to be the first one that philosophers and physicists pursue when investigating a symmetry like time reversal, and I will do so by providing several examples of the philosophical work that realist symmetries can do for us that conventionalist symmetries cannot. I want to reiterate that what follows is not an argument for always adopting the realist stance towards all symmetries; rather, I merely claim that adopting the realist stance towards symmetries is likely to yield more interesting fruit for physicists and philosophers alike, so such a stance is generally the best one to adopt.

2.2.1 Overarching Symmetries and Theory Change

One important role overarching symmetries play within the physical sciences becomes apparent when one considers how such symmetries help us to determine which new theories should replace older, falsified theories. Typically, when an earlier theory has been rejected because of some disconfirming evidence and physicists are searching for a new theory to replace the old theory, physicists don't just start from scratch. When faced with a previously good theory's inadequacies, we typically (rightly) as-

sume that many of the features of the old theory were, in fact, correct, especially if the older theory has been around for a while and provided correct predictions when tested. There must have been *something* that the older theory got right to explain the fact that it was so successful, so if we want our new theories to be strictly better than our old theories, we need to identify those successful components of our old theories and carry them over into our new theories. Successful features of older theories can thus serve as constraints on what good candidates for these theories' replacements should look like. This is, essentially, the insight that may lead philosophers and physicists to appeal to a theoretical virtue like methodological conservatism.[17].

Symmetries are typically features of the world that carry over from older theories to newer theories. Imagine, for instance, a fictitious history of physics in which Newtonian mechanics is falsified after repeatedly observing the behavior of particles traveling close to the speed of light. Under such circumstances, we would certainly be justified in searching for a successor to Newtonian mechanics like special relativity that isn't invariant under Galilean boosts; however, since the observations made can be reproduced exactly in laboratories in Germany and in the United States, and since these observations can be reproduced both now and later, we have no reason to reject the symmetries of spatial and temporal translation under which Newtonian mechanics was invariant. So even if our evidence gives us good reason to reject important features of a previous theory, including some of its symmetries, that same evidence may still uphold many of the symmetries of the old theory. For our new theory to be at least as good as our old theory was, we may need to make our new theory invariant under many of the same symmetries under which the old theory was invariant, but since the old theory may have different variables from our new theory, we must rely on overarching symmetries, not physical symmetries alone, to tell us how to properly constrain our new theory in light of its predecessor.

---

[17]See *Sklar* (1975).

The importance of symmetries in such periods of transition has been the focus of much work in contemporary philosophy of science. Recent work by *Lange* (2007, 2009a,b) emphasizes the fact that in both the context of discovery and the context of justification, physicists employ "meta-laws" to help determine the form the laws must take.[18] Numerous principles or rules can serve as meta-laws on Lange's account, including most notably conservation laws and symmetry principles. The value of meta-laws is that they provide a set of constraints which all of the first-order laws need to obey, and what's more, in many cases they explain why certain laws of nature must hold.

For instance, imagine that we are trying to choose between two theories of the electrostatic interaction between electrons and protons T_1 and T_2, which both account for some set of data we've already collected equally well (allowing for a certain amount of experimental error); T_1 and T_2 are not empirically indistinguishable theories, but at the moment all of the data we have collected does not favor one theory over the other. Both T_1 and T_2 refer to positively and negatively charged particles, and both T_1 and T_2 have only one law that refers to charge. In the case of T_1, that law is Coulomb's law, which requires that $|F| = \frac{kq_1q_2}{r^2}$, where the values of q_i are the charges of two particles, F is a vector field on Newtonian spacetime representing the electrostatic force, and r is the distance between two charged particles. In the case of T_2, however, there is an extra term added to Coulomb's law, which makes this law $|F| = \frac{kq_1q_2}{r^2} + q_1$.[19] Switching positive and negative charges in a world described by T_2, then, would change the electromagnetic force experienced by any two particles since the mass of each particle remains constant while switching positive and negative charges in a world described by T_1 would have no effect since the electromagnetic force

---

[18]It is worth noting that, while the existence of meta-laws is a point in favor of Lange's account of first-order lawhood in that Lange's account of lawhood extends quite naturally to account for meta-laws, it is not clear that Lange's is the only account of lawhood that can do so. The nomic primitivist, for instance, can simply take these meta-laws as primitives.

[19]I also assume that both theories provide us with the convention that the more massive particle counts as particle 1 so that this amended Coulomb's law actually makes sense

between any two charged particles would remain the same. If we assume that the overarching symmetry of "charge swapping" $C : \mathcal{N} \to \mathcal{N}$, where $\mathcal{N}$ is the variable space of both T_1 and T_2, is such that $C(F) = F$, then we find that T_1 has charge-swapping as a symmetry while T_2 does not. Finally, if we take it to be a meta-law that all laws must be invariant under charge-swapping, we have a justification for rejecting T_2 in favor of T_1.

So overarching symmetries as meta-laws provide methodologically conservative constraints that may help us to determine which of the proposed laws we are considering is the best successor to some older physical theory. One special subclass of arguments like the one above are what *Belot* (2003) calls symmetry arguments, which are arguments of essentially the following form[20]: Take some theory $T_1 = \langle E_1, X_1, U_1, N_1 \rangle$. We may find that, relative to some problem we are trying to solve, T_1 is inadequate, and so posit a theory $T_2 = \langle E_2, X_2, U_2, N_2 \rangle$, where none of the equations or variables referred to by T_1 necessarily appears in T_2. We may reasonably ask whether or not all of the (overarching) symmetries under which T_1 is invariant are also symmetries under which T_2 is invariant. If T_2 is not invariant under all of the overarching symmetries under which T_1 is invariant, then we may have a good reason to reject T_2 and search for some other replacement for T_1; after all, in many cases, the symmetries under which T_1 is invariant have the status they have for a reason, namely that they represent physical transformations that leave the world essentially the way we found it. Belot provides a number of examples, specifically from Platonic cosmology, special relativity, classical mechanics, and dust cosmology (among other examples) that show that these sorts of symmetry arguments are ubiquitous in physics, and I will not rehearse Belot's examples again here.

What symmetry arguments and other examples of using symmetries to figure out

---

[20]The following paraphrase of Belot's characterization of symmetry arguments is formulated in terms of my own account of symmetries, and as such, since Belot is committed to a different account of symmetries from my own, what follows is not necessarily a characterization he would agree with.

which theories are the best ones to accept a time show is that we can appeal to symmetries in many cases to determine what form the laws of nature should take. Assume that we have some situation where the newer theory keeps the variable spaces of the older theory but adds new equations to them. A strange consequence of some conventionalist views is that, in such cases, two different operations (or sets of operations) on the *very same* variable space may be treated as one and the same operation. Symmetry arguments will be drastically less effective for the conventionalist since in many cases there will be *some* symmetry under which the new theory is invariant that one may be able to justify identifying with a symmetry of the old theory, and without consistency as a criterion for two physical symmetries' falling under the same overarching symmetry, it is not clear to what the conventionalist can (or should) appeal in order to determine which physical symmetries correspond to the same overarching symmetries.

This worry has been stated rather abstractly, so let's examine a more concrete example. Imagine that we are dealing with a system that we originally assumed could be modeled as something close to the harmonic oscillator. Such a system is governed by a single algebraic equation which sets $e_1 = c \cos x - u$ equal to zero, where c is some constant and u and x are the dependent and independent variables respectively. Since the equation is unchanged by a reflection (call it R) that maps a solution at (x, u) to its mirror image at $(-x, u)$ because cosine is an even function (that is, for any x that satisfies this equation, there's a $-x$ that satisfies it too with the same value of u), this reflection constitutes a symmetry of our system. However, imagine that we now propose changing our theory to one governed by the equation that sets $e_2 = c \cos x + k \sin x - u$ equal to zero, where c and k are constants. This new function e_2, unlike e_1, has an odd function of x in it, meaning that it is no longer invariant under the reflection that takes (x, u) to $(-x, u)$. However, is invariant under the transformation that takes (x, u) to $(-x, u - 2k \sin x)$, and as such this

transformation (call it S) represents a symmetry of the damped harmonic oscillator of our new theory.

Now, suppose that we want to ask ourselves whether the new theory governed by $e_2 = 0$ is invariant under the same symmetry (namely, R) as the old theory. The realist has an easy answer: since both theories appeal to the same variable space and (presumably) x and u are interpretively identical, the overarching symmetry of spatial reflection must behave the same way in the damped theory as in the simple theory, which is to say by mapping the point (x, u) to the point $(-x, u)$. Since the newer theory has an equation that includes an odd function of x, it is not invariant under this transformation, and so the damped theory fails to be invariant under spatial reflection in the way that the original theory was. The mere fact that both e_1 and e_2 are each invariant under *some* symmetry (R and S respectively) does not give us any reason to associate these two physical symmetries with the same overarching symmetry "spatial reflection".

Some conventionalists, however, may be incapable of saying anything given only the information I have provided; after all, R is similar in some ways to S, especially in its transformation of x, and so there may be some sense in which we can call S the counterpart to R and so unite them under the same overarching symmetry[21]. But this seems absurd! The two theories utilize the same variable space and differ only in one term found in the equations of these theories. If one is going to import so much of the essential structure of the older theory into the newer theory, how can one justify changing the form of the symmetry we're interested in? Perhaps not all my readers share this intuition that the information given in my setup of this problematic case should be sufficient to allow us to determine whether the two symmetries in question

---

[21]This is not to say, of course, that the so-called conventionalist could not adopt some condition other than consistency as her criterion for determining which physical symmetries correspond to the same overarching symmetries; my point is only that in the situation described above, I see no principled way that the conventionalist could rule out identifying R and S with the same overarching symmetry without relying on something like the realist's notion of consistency.

correspond to the same overarching symmetry, but what is clear is that symmetry arguments and symmetries as meta-laws will be much less useful as guides to the truth if one adopts the conventionalist standard since the more permissive conventionalist stance allows for more theories to be invariant under the same symmetries than does the realist stance.

Even if cases of theory change where the variable space remains unchanged do not move one to prefer realism to conventionalism, cases where our new theory has a very different variable space from its predecessor may motivate one to embrace realism. After all, if we are to use overarching symmetries as guides to our theorizing that provide us with useful constraints on new theories, these helpful constraints will be more or less helpful insofar as they provide us with only one physical symmetry (or, at most, a few) that corresponds to our overarching symmetry of interest in the new theory. Since realist overarching symmetries will correspond to fewer physical symmetries in the new theory than conventionalist overarching symmetries, it seems like realist symmetries will serve as a more helpful guide for us. Thus, regardless of whether our new theory leads us to abandon the variable space of our old theory, we seem to have good reasons to care more about realist than conventionalist symmetries if we want to use the overarching symmetries of our current theories as constraints on our future physical theories.

To sum up: if we want overarching symmetries to serve as meta-laws that help us determine when we have a good proposal for a successor for some failed physical theory, we need some principled way of determining what the physical symmetries of the new theory should look like given the physical symmetries of the old theory. Realists provide us with such a principle in consistency, and conventionalists do not. I take it, then, that if we're looking for overarching symmetries that will serve as useful meta-laws to guide our scientific inquiry, the ones we should focus our attention on are those realist symmetries that require all physical symmetries corresponding to them

to be consistent with one another, not the conventionalist symmetries whose corresponding physical symmetries may differ wildly in their treatment of interpretively identical variables.

2.2.2 Overarching Symmetries and Ontology

Overarching symmetries don't only serve as guides to which theories we should believe to be true in light of their predecessors; they may also serve as guides to extracting metaphysics from our best available physical theories. In some cases (especially when the symmetries in question are spacetime symmetries) the failure of a particular theory we currently accept to be invariant under an overarching symmetry provides us with a good reason to think that there is some special structure in the world whose existence keeps the theory in question from being invariant under that overarching symmetry. For instance, *Pooley* (2003) takes the fact that quantum field theory is not invariant under parity reversal to indicate that there is some spatial orientation field fixed everywhere by the laws that should be added to our spacetime theory. As stated earlier in this dissertation, *Maudlin* (2007) similarly argues that we should posit some preferred orientation in spacetime since the fundamental laws of physics fail to be invariant under time reversal. For spacetime substantivalists, of course, the failure of a particular theory to be invariant under some symmetry may indicate geometric features of the background spacetime structure. If this sort of reasoning is correct, then there is a sense in which we ought to rely on overarching symmetries (and the features of the physical theories invariant under those symmetries) to determine what fundamental objects must be represented in our physical theories.[22] It seems reasonable to think that an account that requires consistency

---

[22]There is a kind of circularity at work here: a physical theory allows us to derive the symmetries of that theory, and the symmetries of the theory tell us what sorts of physical features we ought to include in our physical theory. But this circularity is better understood as a kind of internal consistency between the equations/variables of a theory and that theory's symmetries. As such, when we know what features of the world our theory represents, this constraint determines the form the symmetries take, while in cases where we know the symmetries of the theory, the symmetries

from overarching symmetries will provide more useful information about the geometric features of spacetime than an account that doesn't since the former makes it clearer when a theory is invariant (or not) under a particular overarching symmetry. As such, there is good reason to think realism is a more natural account of symmetries than conventionalism.

There is a flip-side to the relationship between symmetries and ontology as well, namely the view embraced by *Baker* (2010), *North* (2009), and *Ismael and van Fraassen* (2003) among others, that symmetries act as a guide to surplus structure. Proponents of this view of what I will call physical equivalence (PE) take it to be the case that, when two solutions are related to one another by a symmetry, those two solutions represent the same physical state of affairs, meaning that any difference in the characterization of these two solutions is merely conventional or an artifact of the notation used; that is, the notational differences between two solutions related by a symmetry do not correspond to a deep physical difference between the two while a notational difference between two solutions which are not related by a symmetry does correspond to some deep difference between two physical states. The infamous Leibniz shift arguments provide an example of PE in action, and it is commonplace to use especially empirical symmetries for such arguments; if we were to shift every particle in the universe one meter to the left, no one would be able to tell that anything had changed, so we seem to have good reason to believe that there does not exist any physical structure (like, say, absolute space) that privileges one position in space over another. *Belot* (2013) provides some good reasons to reject PE as applied to all symmetries, though I will not rehearse his argument or examples here; however, even if one rejects PE generally, there is no reason why one can't accept a more specific version of PE that refers to particular overarching symmetries. For instance, though I may not take all solutions related by a symmetry to represent the same state of

---

help us determine what features are present in the physical world if it is as our theory tells us it is.

36

affairs, I may still take all solutions related to one another by a time reversal transformation or a parity reversal transformation to represent the same state of affairs. These overarching symmetries serve as a way to apply something like Occam's razor, allowing us to pare down our theory's metaphysical commitments where notational differences may lead us astray in our metaphysical theorizing.

Using the overarching symmetries to pare down our metaphysical commitments may pull us either towards realism or conventionalism depending on our desiderata for projects in metaphysics. Since conventionalists are frequently able to find symmetries where the realist finds none by flouting consistency, we may feel compelled to take a conventionalist stance if we prefer metaphysical desert landscapes, though note that these landscapes may require us to treat two states characterized by completely different assignments of values to variables as the same if the symmetries appealed to by the conventionalist are permissive enough. The realist, on the other hand, can point to the fact that a criterion like consistency allows us to determine what form a symmetry must take within the context of a particular physical theory given only information about the variables it utilizes and how the transformation in question acts on these variables. Should we believe something like PE for a particular overarching symmetry, realists will have an easier time of recognizing physically equivalent states than will conventionalists who ignore consistency.[23] The pragmatic utility of the realist approach seems more compelling to me than do desert landscapes, and so I take PE to provide us with a good reason to prefer realism (or at least acting like a realist) to conventionalism, but this reason is certainly not decisive.

Yet the relationship between symmetries and the ontology of physical theories

---

[23]Note that it is possible to have your cake and eat it too here; one can adopt the stance of the conventionalist who acts like a realist when deciding which symmetries to examine. This conventionalist doesn't take consistency to be the true gauge of what unites physical symmetries under a single overarching symmetry but believes that overarching symmetries that satisfy it are the ones we should examine first and think are the best candidates for the sorts of overarching symmetries we care about. Again, I have no quarrel with such conventionalists; my task is only to argue that we ought to act like realists when searching for overarching symmetries as these peculiar conventionalists do.

gives us another reason to think that, even if we are conventionalists, we should focus our attention on those symmetries that transform the same objects and relations consistently across physical theories. If we do think that symmetries can help us determine the ontological commitments of our physical theories, and if we think that all of the physical theories are, in fact, representations of the same objective world, then it would seem desirable to have the ontologies identified by different physical theories line up nicely with one another. For instance, if one theory is committed to the existence of a preferred inertial reference frame (as some formulations of non-relativistic quantum mechanics are) and another theory is committed to the physical equivalence of all inertial reference frames (in the way that special relativity is if we assert something like PE for velocity boosts), we obviously have good reason to think that one of these theories is correct and one is wrong since only one can accurately represent the way the world actually is.

If we want to use symmetries as a guide to surplus structures or use asymmetries to justify the existence of certain structures, then it would seem that the symmetries even the conventionalist would most want to focus on should be those that can be extended consistently beyond the scope of one particular theory to act on all variables that represent the same physical property consistently. For instance, take two theories T_1 and T_2 which are both invariant under Galilean velocity boosts (they could be, for example, classical mechanics and classical electrostatics). If we have good reason to think that both T_1 and T_2 are true, then we have good reason to think that any two inertial frames are physically equivalent to one another, and thus that any additional structure posited to distinguish the two frames (such as, for instance, some absolute velocity) doesn't really exist. If T_2 is invariant under Lorentz boosts but not Galilean boosts, however, our ontological commitments based on accepting both T_1 and T_2 would be quite different (now take T_2 to be, for example, special relativity); after all, a world in which I'm traveling inertially at five times the speed of light would be

allowed by T_1 but not by T_2 since superluminal velocities are forbidden by the laws of T_2. In such cases, we have good reason to believe that either T_1 or T_2 is false since there can be no single theory that respects the symmetries and asymmetries of both physical theories simultaneously[24].

If we want to know what we really ought to believe exists, we need to look at the commitments of *all* of our best physical theories, not just one or two. If we want to get a consistent picture of what the world is like from these theories, and if we think that symmetries can tell us something about the ontological commitments of our physical theories, then even the conventionalist needs to accept the fact that overarching symmetries that act consistently on the fundamental objects and relations of lots of different physical theories will be more interesting than overarching symmetries that take entirely different forms depending on the theory being examined. So, one could say that, despite her conventionalism, if a conventionalist has the inclination to draw a connection between physical symmetries and ontology, then she should be prepared to act as a realist in certain situations.

2.2.3   Unification

As discussed in the previous section, the fact that we take all physical theories to represent the nature of one and the same world gives even the conventionalist reason to act like a realist in many cases by focusing her attention on overarching symmetries that act consistently across a number of different physical theories before examining symmetries whose form varies from theory to theory. The fact that our best physical

---

[24]Here, and later on in the paper, one may worry about the clear inconsistencies among our best physical theories, most notably general relativity and quantum mechanics. Such inconsistencies typically lead to one of two conclusions: either the physical world is not the sort of thing that can be completely modeled by completely consistent physical theories, or else the world is capable of being so modeled and at least one of the best physical theories we are currently working with is either simply false or, perhaps more charitably, incomplete and misleading. Though my sympathies lie with the latter of these two camps, I will not try to defend this position here. All I mean to argue is that those who think fundamental physical theories can and should be made consistent with one another ought to adopt the realist strategy when seeking out overarching symmetries.

theories tend to bump up against one another and interact with one another gives us further reason to look for realist symmetries first as well. We infrequently think that any single theory we've discovered so far is sufficient to account for all of the phenomena we'd like our physical theories to be able to account for, so we rely on a cast of different physical theories to provide us with an accurate representation of what the world is really like and accurate predictions of what experimental outcomes we will observe.

There are some, however, who still hold out for a "Grand Unified Theory of Everything" (GUTE) that will eventually replace the multiplicity of theories we rely on to do science's explanatory and predictive work. The hope is something like the following: there have been several cases in the history of science where older theories were replaced by more unifying theories. For instance, Newtonian mechanics was capable of predicting and explaining both celestial and terrestrial motion, a feat that had not been reproduced by any empirically adequate prior physical theory. Maxwell's theory of electromagnetism provided an unexpected unified theory of a electricity, magnetism, and optics. The electroweak theory provides a single theory to account for both the weak and electromagnetic forces. The hope of the scientists (and philosophers) who anticipate some future GUTE is that, at some future point, physics will provide suitable accounts of all forces (including the strong nuclear force and gravity) under the same theory. We may not know what form the GUTE will take, but we can hope that we will discover such a theory and prepare our current theories to make them as amenable to a future GUTE as possible.

It is perhaps with a GUTE in mind that searching for realist symmetries first seems most appealing. The advocate of a future GUTE has the hope that, at some point in the future, there will be a single theory that accounts for the disparate phenomena that we today rely on a number of different scientific theories to account for. The GUTE advocate, then, is looking for clues as to the form the GUTE will

take, and in this regard the sorts of symmetries the realist searches for provide more helpful information than do symmetries of the sort the conventionalist allows because they pick out fewer candidate symmetries. If all of our best current theories are invariant under some set of symmetry transformations, then we have good reason to think that the GUTE will be invariant under such symmetry transformations as well, and if all of our best physical theories are not invariant under a particular symmetry transformation, then we may have no good reason to believe that the GUTE we hope for will be invariant under this particular symmetry either (assuming, of course, that the GUTE utilizes many of the same parameters that our current best physical theories use).[25]

Realist symmetries here will be more useful because the consistency constraint ensures that physical symmetries united under the same overarching symmetry will transform the same variables in the same ways. So, for instance, if all of our best theories are invariant under a realist spatial translation, then, when positing a GUTE, we have good reason to think that it will be invariant under a spatial translation that behaves in our new theory just the way it behaved in our old theories. Realist symmetries thus provide the GUTE advocate with potentially helpful information about what form the GUTE may take, assuming of course that there is something basic that our current scientific theories have gotten right about the world, namely what symmetry transformations physical theories ought to be invariant under. Again, I do not mean to don the mantle of the GUTE advocate myself; it simply seems a reasonable view to take about the progress of science, and it's one that seems to lend itself to adopting the realist strategy over the conventionalist strategy. My point is only that the consistent symmetries, in this situation, may provide us with more information than the symmetries the conventionalist allows. Whether one be realist

---

[25]My point here is, in a sense, an extension of my point about realist symmetries being the most useful ones for helping us determine the form of successor theories since the GUTE is supposed to be the successor for all of our current physical theories.

or conventionalist in one's understanding of symmetries like time reversal, it seems the best course of action, in most cases, to look for consistent symmetries first to glean as much information as possible from these symmetries about the world our scientific theories purport to represent.

2.3   Conclusion

Assuming that my approach to physical symmetries is correct, we are left with at least two ways to proceed: when searching for overarching symmetries, we can follow the realist and search for symmetries that transform the same objects and relations consistently across physical theories (for only such transformations deserve to be labeled overarching symmetries), or we can follow the conventionalist and also search for symmetries that may not transform the same objects and relations consistently across physical theories but instead preserve some other important feature of what it means to be a time reversal operation, say, or a parity reversal operation. As the preceding sections show, conducting the realist's narrow search is more likely to yield the kinds of results that both scientists and philosophers are likely to find pragmatically useful, so it may be in the best interest of even the conventionalist to carry out the realist's narrow search first before conducting the broader search.

The moral of this story is one that needs to be taken to heart in the discussion of time reversal in particular. What my analysis suggests is that, when looking for a suitable candidate for a time reversal operator that may have interesting philosophical or physical consequences (e.g. one that may be able to guide our ontology, or one that may be suggestive of the form the GUTE may take), we should not broaden our search to include inconsistent symmetries that intuitively seem to play the role of a time reversal operator within the context of a specific theory; rather, we should narrow our search and count as candidate time reversal transformations only those which act consistently on the same variables across a number of different physical

theories. In the coming chapters, I will use this approach to provide a new account of time reversal and derive from it a number of interesting consequences.

CHAPTER III

Background

3.1 What is time reversal?

The following two questions about time reversal are intimately related to one another:

1. What does the time reversal operator look like (or, equivalently, how do physical properties change under time reversal)?

2. Which physical theories are time reversal invariant?

In the philosophical literature on time reversal, authors frequently attempt to answer one of these questions by assuming an answer to the other question and using the details from this first solution to argue for a solution to the original question. Why they do so seems obvious: if one knows how the time reversal operator acts on physical states, it is relatively easy to conjure up a time reversal operator in the context of a particular physical theory and then check to see whether this time reversal operator maps solutions of this physical theory to solutions. Conversely, if one assumes from the beginning that a particular theory is time reversal invariant, one can utilize the mathematical structures of the theory (e.g. the symmetries under which the theory's differential equations are invariant) to determine what properties a time reversal

44

operator should satisfy. I will call accounts of time reversal that assume an answer to 1 and use this response to generate an answer to 2 "intuitive" and accounts of time reversal that assume an answer to 2 and use this response to generate an answer to 1 "theory-relative".

In this chapter I examine several accounts of time reversal all of which are either intuitive or theory-relative. In the first section I consider the work of *Horwich* (1987), *Albert* (2000), *Malament* (2004), and *Arntzenius and Greaves* (2009), all of whom provide intuitive accounts of time reversal. In the second section I consider the work of *Roberts* (2010a,b) and some claims made by the "textbook account" considered by Arntzenius and Greaves, both of which are theory-relative accounts of time reversal. Finally, I conclude with some thoughts about the general shortcomings of both intuitive and theory-relative accounts and sketch a third approach that may provide a better account of time reversal than any in the literature to date.

3.2   Intuitive accounts of time reversal

Intuitive accounts of time reversal may differ from one another, but they all employ the same general strategy for determining whether particular theories are time reversal invariant, which runs basically as follows:

1. Begin with familiar physical properties (such as position and perhaps velocity) of whose behavior under the time reversal operator we have an intuitive grasp.

2. Utilize some formal relations provided for you by the theory in question to determine how the values of other fundamental physical properties of the theory transform under time reversal.

3. Check to see if any solution of the theory is mapped to a solution by

the putative time reversal procedure. If so, then the theory is time

reversal invariant. If not, then the theory is not time reversal invariant.

A general procedure for laying out a taxonomy of intuitive accounts of time reversal, then, can be given by providing the following information about each account: 1) the intuitions which drive the particular characterization of time reversal needed for the author's account, 2) the consequences of this characterization of time reversal for the transformation of the fundamental properties of a physical theory under time reversal, and 3) the verdict the account delivers concerning the time reversal invariance of particular theories. I will now characterize four different intuitive accounts of time reversal to examine how such accounts of time reversal work, what their strengths are, and where their weaknesses lie.

3.2.1 Horwich

I begin my inquiry into intuitive accounts of time reversal with chapter 3 of *Horwich* (1987). Horwich, the first to my knowledge to provide a rigorous philosophical account of time reversal in general and an intuitive account of time reversal in particular, begins by pointing out that it is not enough for a time reversal operator to simply reverse the order of physical states; that is, it is not sufficient to characterize $S_n...S_1$, a simple inversion of some sequence of physical states $S_1...S_n$, as the result of every time reversal operation. Though Horwich agrees that intuitively time reversal should at the *very least* reverse the temporal ordering of instantaneous physical states, something more is needed for a full account of time reversal. Thus, some operator (which I shall call T, using my own notation instead of Horwich's) should be employed to transform the individual states into their time-reversed counterparts. Horwich claims that the true time-reversed version of the sequence $S_1...S_n$ is $T(S_n)...T(S_1)$.

Having appealed to the time reversal operator T above, Horwich now owes us an account of how that operator acts on physical states. Horwich begins his rather quick

analysis of T by drawing the distinction between what he calls "basic" properties, which are (presumably monadic) properties held by instantaneous physical states, and "non-basic" properties, which are relations between numerous instantaneous states across time. Spatial positions and time, for instance, count as basic properties in Horwich's account while velocity is non-basic. Horwich defines the operation of T on the instantaneous state $P_1 = \langle \mathbf{x_n}, t \rangle$, where $\mathbf{x_n}$ is the position of the relevant instantaneous physical state and t is temporal coordinate, as follows:[1]

$$T(\mathbf{x_n}) = \mathbf{x_n} \tag{3.1}$$

$$T(t) = -t \tag{3.2}$$

These are the only basic properties that Horwich identifies, and so Horwich's account will, of course, be of limited use; however, it does provide enough information for us to understand how the operator T acts on the velocity $\mathbf{v}$ of P_1. Since the velocity is the time derivative of position, and since T flips the sign of time but not of position, it follows that:

$$T(\mathbf{v}) = -\mathbf{v}. \tag{3.3}$$

Horwich does not explicitly discuss the ways in which other non-basic properties transform under T, but he does hint that the magnetic field $\mathbf{B}$ should be treated similarly to $\mathbf{v}$ since flipping the direction in which electrons flow in a wire would

---

[1]For the purposes of this paper, I will use the notation $T(x) = x'$ to represent the projection of the transformation T onto the subspace characterized by the variable x. Less technically, when I use the notation $T(x) = x$, I isolate the effect of the time reversal operator to those components of the state Ψ that depend on certain variables (in this case, x) while ignoring the effect of time reversal on all other components of Ψ. This is also what I mean when I say that T is "acting on" a property or variable x.

likewise flip the sign of the magnetic field these electrons induce:

$$T(\mathbf{B}) = -\mathbf{B}. \tag{3.4}$$

Note that Horwich's maneuver here seems at odds with my interpretation of him as giving an intuitive account of time reversal because it essentially assumes the time-reversal invariance of Ampere's circuital law (with Maxwell's correction):

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j}; \tag{3.5}$$

that is, the only reason why flipping the directions of the electrons in a loop of wire would flip the sign of the magnetic field induced by those electrons would be if the above equation were time reversal invariant. So Horwich's intuitions about time and velocity alone are not sufficient to justify his account of the action of the magnetic field under a time reversal transformation. It is worth noting Horwich's treatment of $\mathbf{B}$ under T and his justification for it here because these will be points of contention between Horwich and Albert, as I shall discuss in the next section.

Finally, given the above descriptions of how the time reversal operator T acts on instantaneous physical states, what can we conclude about the time reversal invariance of particular physical theories? Horwich does not draw any conclusions of his own on this matter in chapter 3 of his book, but his suggestion that the magnetic field flips its sign under time reversal allows us to extrapolate Horwich's views on the time reversal invariance of classical electromagnetism. The fundamental equations in classical electromagnetism are Maxwell's:

$$\nabla \cdot \mathbf{E} = \rho \tag{3.6}$$

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j} \tag{3.7}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{3.8}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{3.9}$$

$$\tag{3.10}$$

and the Lorentz force law:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{3.11}$$

As it stands, Horwich does not provide us with enough information to determine whether or not these equations are invariant under time reversal because he does not state how T operates on the electric field $\mathbf{E}$, the current $\mathbf{j}$, or the electromagnetic force $\mathbf{F}$. However, it seems like Horwich would likely support the following characterizations of the action of T on these variables:

$$T(\mathbf{E}) = \mathbf{E} \tag{3.12}$$

$$T(\mathbf{j}) = -\mathbf{j} \tag{3.13}$$

$$T(\mathbf{F}) = \mathbf{F} \tag{3.14}$$

The transformation of these properties under the time reversal operator can be justified as follows: first, since $\mathbf{F} = m\mathbf{a} = m\frac{d^2\mathbf{x}}{dt^2}$ by Newton's second law, it seems that $\mathbf{F}$ should only flip signs if m flips its sign under T since $\mathbf{x}$ does not flip signs under T

and the sign flip of t under T does not matter because we are dealing with a second derivative.[2] But there seems to be no good reason to think that time reversal should have any effect on m, and so our intuitions suggest that $T(m) = m$. Just as mass seems to be a property that has nothing to do with time, the electric field $\mathbf{E}$ likewise seems a property that time reversal should not affect, and so Horwich seems likely to advocate the transformation of $\mathbf{E}$ suggested above. Finally, since the current $\mathbf{j}$ is a quantity that involves a single time derivative, it should intuitively flip signs under time reversal. All three of the above extrapolations, then, seem fairly reasonable given Horwich's view, and as we'll see shortly, Albert agrees with Horwich on these points.

All that remains now is to check and see if T, as I have characterized it, leaves Maxwell's equations and the Lorentz force law invariant. As a matter of fact, the operator T does transform physical states in such a way that if some state P_1 has properties that obey Maxwell's equations and the Lorentz force law, the time-reversed state $T(P_1)$ has properties that obey Maxwell's equations and the Lorentz force law too, and so classical electromagnetism appears to be time reversal invariant.

3.2.2 Albert

Over a decade after Horwich's book was published, *Albert* (2000), in the first chapter of his book, provides an account of time reversal that, for the most part, agrees with Horwich's in spirit even as it differs from Horwich's account on details pertaining to classical electromagnetism. In many ways Albert's general project follows Horwich's closely. Like Horwich, Albert divides physical properties into basic properties and non-basic properties, and he agrees with Horwich's assessment that

---

[2]I assume here that $\mathbf{F} = m\mathbf{a}$ is essentially definitional and not a statement of some sort of substantive law. Definitions, I take it, are assumed to be time reversal invariant and cannot fail to be so while substantive physical laws can fail to be time reversal invariant. If we take Newton's second law here to be a substantive physical law, then we have to make an assumption about the time reversibility of this law to determine how T acts on $\mathbf{F}$, just as Horwich did with $\mathbf{B}$.

the time reversal operator T acts in the following way on the following basic and non-basic physical properties:

$$T(\mathbf{x_n}) = \mathbf{x_n} \tag{3.15}$$

$$T(t) = -t \tag{3.16}$$

$$T(\mathbf{v}) = -\mathbf{v} \tag{3.17}$$

However, Albert's intuitions differ from Horwich's on a few important points. The difference between Horwich and Albert, in a sense, boils down to the fact that Horwich assumes to time reversal invariance of Ampere's circuital law, either because he takes it to be definitional or just "obviously" time reversal invariant, while Albert takes Ampere's circuital law to be just another of the laws of classical electromagnetism whose invariance under time reversal is up for grabs. Albert suggests that time reversal ought to simply reverse the temporal ordering of physical states but leave all of the fundamental quantities invariant.[3] That is, according to Albert, T should never flip the sign of a property's value unless that property is either a temporal coordinate or a non-basic property defined as a time-derivative of some other, more basic property whose values do not flip sign under time reversal. Despite Horwich's motivations for flipping the sign of the magnetic field under time reversal, Albert claims that the magnetic field, since it is a basic property and not explicitly the time-derivative of some more fundamental property, should not flip sign under time reversal. Thus, Albert defines the operation of T on the basic and non-basic properties of classical electromagnetism as follows:

---

[3]On this point, *Callender* (2000) agrees with Albert, though Callender's own analysis focuses on the time reversal invariance of non-relativistic quantum mechanics as opposed to electromagnetism.

$$T(\mathbf{B}) = \mathbf{B} \tag{3.18}$$

$$T(\mathbf{E}) = \mathbf{E} \tag{3.19}$$

$$T(\mathbf{j}) = -\mathbf{j} \tag{3.20}$$

$$T(\mathbf{F}) = \mathbf{F} \tag{3.21}$$

Albert's version of T thus yields the same transformations as Horwich's with the sole exception of the magnetic field; however, this change is sufficient to render classical electromagnetism, which was invariant under Horwich's proposed time reversal operator, non-invariant under Albert's. To see this, consider again Ampere's circuital law, $\nabla \times \mathbf{B} = \dfrac{\partial \mathbf{E}}{\partial t} + \mathbf{j}$. The right-hand side of this equation flips signs under time reversal while the left-hand side does not. There are no non-zero values of $\mathbf{B}$ which would satisfy both Maxwell's second equation and its time-reversed counterpart, so classical electromagnetism is not, according to Albert, time reversal invariant.

3.2.3   Malament

Up to this point I have ignored the specific ways in which intuitive accounts typically interpret or understand how the time reversal operator acts on states; however, the issue of interpretation cannot be avoided when discussing the distinctive features of the account given by *Malament* (2004), so a little bit of backtracking is in order. As *North* (2008) and *Arntzenius and Greaves* (2009) point out, time reversal is typically interpreted or understood in one of two essentially equivalent ways. The first sort of time reversal, active time reversal, involves a temporal flipping of physical states (as characterized by the contents of spacetime and their fundamental physical properties) while leaving the coordinate system and background spacetime of these states unaltered. One can think of such time reversals as making the necessary changes to the

progression of physical states such that, if you observed the time-reversed sequence of a movie, for instance, the movie would appear to be running backwards. The second sort of time reversal, passive time reversal, involves simply a reassignment of coordinates without changing any intrinsic features of the physical states in question. One can perhaps think of the passive time reversal of some sequence of events as the way in which someone moving backwards in time would view the original sequence. Essentially, these two physical interpretations of the time reversal operator amount to the same thing: whether one keeps the coordinates fixed while time-reversing the physical states or keeps the physical states fixed while time-reversing the coordinates, the fundamental characterization of the time reversal operator should remain the same.

Malament proposes a new physical interpretation of time reversal. Rather than supposing that time reversal acts on the physical states alone or coordinates alone, Malament takes the time reversal operator to invert the temporal orientation of the background spacetime structure. This interpretation, which he calls "geometric", is more general than the passive interpretation since it applies to coordinate-independent models, and Malament claims that, unlike active and passive interpretations of time reversal, his geometric interpretation applies to curved spacetimes. So it seems that of the intuitive accounts considered thus far, Malament's is the best-suited to deliver verdicts on the notion of time reversal in our best currently available physical theories.[4]

So, what verdict does Malament's geometric interpretation deliver in the case of classical electromagnetism? Malament begins by providing us with the entities and properties he takes to be basic. In the context of classical electromagnetism, these are charged particles and fields, in particular the electromagnetic field F which Malament takes to be a map from the pair $\langle L, q \rangle$ at some point p, where L is the line tangent to

---

[4]This point is also made by *North* (2008).

some particle's worldline at p and q is the charge of the particle at p, to a four-vector indicating the direction in which the test particle characterized by $\langle L, q \rangle$ would move if placed at p. Malament notes that some information is missing in his characterization of F because, for any timelike worldline, there are two possibilities for the unit tangent vector at a given point: one which points towards the past, and one which points towards the future. Since there is nothing in Malament's characterization so far to suggest that L should be future-directed instead of past-directed, Malament relies on the temporal orientation of the background spacetime to determine the direction in which L points. So F, according to Malament, takes the pair of objects $\langle L, q \rangle$ to some four-vector given a specific background temporal orientation.[5]

Geometric time reversal involves keeping the fundamental quantities the same while flipping the temporal orientation, so the operator T should have no effect on the position or charge of a particle. Thus, according to Malament's view:

$$T(\mathbf{x_n}) = \mathbf{x_n} \tag{3.22}$$

$$T(q) = q \tag{3.23}$$

However, Malament agrees with both Horwich and Albert that velocities flip sign under time reversal:

$$T(\mathbf{v}) = -\mathbf{v} \tag{3.24}$$

This follows from the fact that velocities are future-directed four-vectors that lie

---

[5]It is worth noting that a similar convention for an orientation in spacetime could be picked out by positing something like the local orientation field considered by *Pooley* (2003). I will discuss similarities and differences between orientation fields and inherent temporal orientations in background spacetime structure in the final chapter of this dissertation.

tangent to worldlines. Thus, the role of temporal orientation in determining velocity requires a velocity sign flip under geometric time reversal. Like velocity, the field F requires a temporal orientation for its mapping to work. Intuitively, L and q do not flip signs under geometric time reversal since the fundamental quantities of a physical theory remain untouched by a geometric time reversal; however, the four-vector that results from the mapping ought to flip its sign under time reversal because of the temporal orientation introduced to F as discussed above. Thus, we should expect the electromagnetic field, represented by the Maxwell-Faraday tensor F^{ab}, to transform as follows:

$$T(F^{ab}) = -F^{ab} \tag{3.25}$$

This does not, however, tell us how $\mathbf{E}$ and $\mathbf{B}$ transform, so it is hard to compare Malament's account to Horwich's and Albert's accounts without saying a bit more about the relationship between $\mathbf{E}$, $\mathbf{B}$, and F^{ab}. Skimming over the technical details of Malament's account, he defines the electric field $\mathbf{E}$ as the product of the tensor F^{ab} and a "frame", which is a future-directed time-like vector field and thus a physical quantity whose value flips signs under geometric time reversal. Since both F^{ab} and the frame flip signs under time reversal, $\mathbf{E}$ itself is not affected by the time reversal operator. The magnetic field $\mathbf{B}$, however, is determined to be proportional to the product of F^{ab}, the frame, and a volume element representing an antisymmetric tensor field whose values also flip sign under geometric time reversal. Thus, $\mathbf{B}$ itself flips sign under geometric time reversal. So, like Horwich, Malament takes T to operate in the following way on the following quantities of interest in classical electromagnetism:

$$T(\mathbf{B}) = -\mathbf{B} \tag{3.26}$$

$$T(\mathbf{E}) = \mathbf{E} \tag{3.27}$$

$$T(\mathbf{j}) = -\mathbf{j} \tag{3.28}$$

$$T(\mathbf{F}) = \mathbf{F} \tag{3.29}$$

As previously discussed, Maxwell's equations and the Lorentz force law all still hold under these transformations, and thus classical electromagnetism, according to Malament, is time reversal invariant.

3.2.4 Arntzenius and Greaves's "Feynman" account

The last intuitive account of time reversal I consider comes from *Arntzenius and Greaves* (2009). Arntzenius and Greaves agree with Malament's general strategy but also provide a new strategy (which, they claim, is equivalent to Malament's) inspired by Feynman's claim that "antiparticles are just particles moving backwards in time". Essentially, the Arntzenius-Greaves (henceforth AG) program begins by building in an intrinsic temporal orientation to the tangent worldlines L utilized by Malament in his analysis. If worldlines are intrinsically-directed, as Malament claims they are not, then L flips signs under time reversal, which means that the field F is no longer only defined relative to a specific temporal orientation, so:

$$T(F^{ab}) = F^{ab}. \tag{3.30}$$

If one follows Malament's procedure for deriving the behavior of $\mathbf{E}$ and $\mathbf{B}$ under time reversal from the behavior of F^{ab} under time reversal as described in the previous

section, then one finds that $\mathbf{E}$ is defined in terms of the product of a frame which flips signs under time reversal and a F^{ab} which does not, meaning that $\mathbf{E}$ now flips signs under time reversal. Likewise, $\mathbf{B}$ now does not flip its sign under time reversal.

A few more important quantities should also be discussed. If one is to take Feynman at his word on the idea that anti-particles are simply particles moving backwards in time, one must saddle Feynman with the view that the time reversal operator flips the charge of the particle instead of just leaving it invariant; after all, how else could electrons and positrons have opposite charges? Perhaps more curiously, velocity and velocity-dependent quantities like j now do not flip signs under time reversal because their temporal directness is treated as basic instead of parasitic on the underlying temporal orientation of spacetime. Thus, the "Feynman proposal" of AG defines the operation of T as follows:

$$T(\mathbf{x_n}) = \mathbf{x_n} \tag{3.31}$$

$$T(q) = -q \tag{3.32}$$

$$T(t) = -t \tag{3.33}$$

$$T(\mathbf{v}) = \mathbf{v} \tag{3.34}$$

$$T(\mathbf{B}) = \mathbf{B} \tag{3.35}$$

$$T(\mathbf{E}) = -\mathbf{E} \tag{3.36}$$

$$T(\mathbf{j}) = \mathbf{j} \tag{3.37}$$

$$T(\mathbf{F}) = \mathbf{F} \tag{3.38}$$

If T is defined in this way, it turns out that any set of values $\langle T(\mathbf{x_n}), T(q), T(t), T(\mathbf{v}), T(\mathbf{B}), T(\mathbf{E}), T(\mathbf{j}), T(\mathbf{F}) \rangle$ will satisfy Maxwell's equations and the Lorentz force law just in case the set of values $\langle \mathbf{x_n}, q, t, \mathbf{v}, \mathbf{B}, \mathbf{E}, \mathbf{j}, \mathbf{F} \rangle$ does; thus, classical electromagnetism is invariant under the "Feynman" time reversal operator just as it was under

Malament's time reversal operator even though these two operators act very differently on physical states.[6]

3.3 Theory-relative accounts of time reversal

The alternative to the intuitive accounts of time reversal examined above are what I have called theory-relative accounts of time reversal. These accounts proceed in two seemingly simple but technically complex steps by 1) singling out a particular physical theory which intuitively seems time reversal invariant and 2) determining what features the time reversal operator would need to have in order for the given theory to be time reversal invariant. This approach has been fairly unpopular in the philosophical literature on time reversal, so much so that there are only two truly theory-relative accounts discussed in the literature at the moment: one briefly discussed in Arntzenius and Greaves's work and one provided by *Roberts* (2010a,b).

3.3.1 Arntzenius and Greaves's "textbook" account

Arntzenius and Greaves (2009) discuss what they call the "textbook account" of time reversal at the beginning of their paper on the time reversal invariance of electromagnetism as a foil for the accounts that follow. Very quickly, Arntzenius and Greaves claim that most classical electromagnetism textbooks argue first that electromagnetism ought to be time reversal invariant and, from this assumption, derive the following properties of the time reversal operator:

---

[6]It is worth noting, having discussed AG's "Feynman" account and Malament's geometric account, that AG believe that the difference between these two accounts is merely conventional. They provide a "structuralist" ontology according to which certain choices of convention allow one to derive the basic features invoked by Malament's account or the basic features invoked in "Feynman's" account. There may, then, be no genuine, non-conventional difference between these two accounts despite appearances to the contrary.

$$T(\mathbf{x_n}) = \mathbf{x_n} \tag{3.39}$$

$$T(q) = q \tag{3.40}$$

$$T(t) = -t \tag{3.41}$$

$$T(\mathbf{v}) = -\mathbf{v} \tag{3.42}$$

$$T(\mathbf{B}) = -\mathbf{B} \tag{3.43}$$

$$T(\mathbf{E}) = \mathbf{E} \tag{3.44}$$

$$T(\mathbf{j}) = -\mathbf{j} \tag{3.45}$$

$$T(\mathbf{F}) = \mathbf{F} \tag{3.46}$$

Clearly, the above operator is the same as the one invoked by Horwich and Malament, at least in its operation on these particular properties. Two points are worth noting here that I will come back to in the final section of this chapter as I critique both the intuitive and theory-relative accounts of time reversal. First, note that the "textbook account" is able to deliver the features of the time reversal operator T without making any assumptions about the fundamental properties of the physical theory or which interpretation of time reversal (active, passive, geometrical) was best. Given this ontological or interpretive agnosticism, the textbook account may seem alluring to many philosophers of physics. However, my second point suggests a problem with the textbook account, which is that the operator it delivers in this case isn't the only candidate for T that keeps electromagnetism time reversal invariant. As discussed in the last section, the "Feynman" account provides a time reversal operator that acts very differently from the one provided by the textbook account, and yet classical electromagnetism is still time reversal invariant under this strange-looking operator. One may rightly worry, then, that theory-relative accounts of time reversal do not provide enough information to pick out a unique form for the time reversal operator

to take within the context of a particular physical theory. Both of these worries will be raised again and discussed in more detail in the next section.

3.3.2 Roberts

Roberts (2010a,b) also provides a theory-relative account of time reversal, primarily in the context of quantum mechanics. By assuming that quantum mechanics is time reversal invariant, Roberts is able to derive a number of properties of the quantum mechanical time reversal operator, including 1) its antiunitarity, 2) its commutation and anticommutation relations with the position and momentum operators, 3) its commutation relations with spin operators, and 4) its operation on position and momentum values in classical mechanics. Since the purpose of this section of the paper is to deliver a comparison among many different accounts of time reversal, and since the other authors I have considered do not deliver verdicts on the time reversability of quantum mechanics, I will focus primarily on Roberts's fourth result instead of his first three.

After Roberts determines the general time reversal operator for quantum mechanics, he ventures into the classical realm to check and see how it operates on classical position and momentum. The following are his strange but interesting results: the time reversal operator in classical mechanics is either T_1 or T_2, where

$$T_1(\mathbf{x_n}) = \mathbf{x_n} \tag{3.47}$$

$$T_1(\mathbf{v}) = -\mathbf{v} \tag{3.48}$$

and

$$T_2(\mathbf{x_n}) = \mathbf{x_0} - \mathbf{x_n} \tag{3.49}$$

$$T_2(\mathbf{v}) = \mathbf{v} \tag{3.50}$$

for some fixed $\mathbf{x_0}$. Either T_1 or T_2, Roberts claims, serves as a suitable time reversal operator for classical mechanics, strange as that may sound since T_2 seems more like a space *and* time reversal operator, flipping $\mathbf{x_n}$ around some constant $\mathbf{x_0}$ (which, according to Roberts, represents our freedom to choose an axis about which to flip space), but Roberts's account gives us no reason to prefer T_1 to T_2 as the best time reversal operator for classical mechanics. Again, more will be said about this feature of Roberts's account in the next section.

3.3.3 A third way?

Though both intuitive and theory-relative accounts have their advantages, both of these general strategies seem unappealing for a number of reasons. In this section, I first discuss the problems facing each account and then suggest a potential reconciliation of the two accounts as well as the kinds of problems that such a third account could solve.

3.3.4 Problems for intuitive and theory-relative accounts

The most popular approach to time reversal seems to be the intuitive approach, and sophisticated accounts such as Malament's and AG are certainly interesting and allow for exciting work on time reversal; however, the intuitive approach has problems generally, and many particular intuitive accounts run into idiosyncratic problems. I begin with the specific problems and then work outwards toward more general issues.

One worry about Malament's and Arntzenius and Greaves's projects is that both

assume an affirmative answer to the question "Is our spacetime temporally 'handed'?". One of the uses for a time reversal operator may be to help us answer the questions of 1) whether time is handed or not, and 2) if time is handed, why it is so. Horwich is moved by these questions to provide his own analysis of time reversal. If one is committed to something like the spacetime symmetry principles advocated by *Earman* (1989), one of which is states that all spacetime symmetries are dynamical symmetries, then perhaps we should conclude that a time reversal asymmetry in one of our dynamical theories should indicate a temporal handedness in spacetime. In short, the existing literature gives us a good reason to believe that an analysis of time reversal may help to answer the question of whether or not time is handed;[7] however, because Malament and AG assume that time is handed in order to formulate their accounts of time reversal, these accounts are incapable of providing evidence for the handedness of time, so those interested in such questions need to look elsewhere for accounts of time reversal that will further their projects.

Perhaps a more difficult challenge to intuitive projects in general is the one that the AG program suggests, namely that all intuitive accounts rely on some basic assumptions about the fundamental ontology of a certain physical theory. The division of properties into basic and non-basic certainly seems a metaphysically weighty endeavor with potentially enormous consequences for one's preferred intuitive account of time reversal. One might prefer, as I suggested in my discussion of the "textbook account", some account of time reversal that does not require such assumptions. What is more, while it may be easy to determine what sorts of elements are best admitted into one's ontology in theories such as classical electromagnetism and classical mechanics, more recent physical theories such as string theory and quantum chromodynamics may be such that even our best physical intuitions do not lead straightforwardly to one

---

[7]I will argue in the final chapter of the dissertation that this argument is not nearly as straightforward as it seems to be, but this does not mean that that a well-motivated account of time reversal has *no* bearing on questions about the handedness of time.

particular ontology or another.[8] With such dubious metaphysics, one wonders if the success of accounts like Malament's and AG's "Feynman" proposal can be repeated in the context of these new theories, and if not, it again seems like there are many cases where our intuitions should not be our guide for determining what the time reversal operator looks like.

Roberts provides his own challenge to these intuitive accounts, arguing that, while our intuitions may be good guides to determining how properties like classical position and momentum transform when time-reversed, our intuitions may mislead or abandon us when we're faced with strange new physical theories and odd physical properties. We may know intuitively how velocity should transform under time reversal, for instance, but what about lepton number, isospin, quark color or flavor, or any of the other numerous properties that we find in our best available physical theories? What's more, should we consider these properties to be "basic" in Horwich's and Albert's sense? With few, weak, or unreliable intuitions to guide us, it seems likely that we will miss something when providing an intuitive account of time reversal, so we have good reason to turn to some other approach in these cases.

Finally, when compared to the theory-relative approach, the intuitive accounts seem to rely on the problematic assumption that we know, *a priori*, what the time reversal operator looks like in the context of a given physical theory. As my previous criticisms have suggested, our intuitions do not seem to be excellent guides for what we should believe to be true in our fundamental physical theories, so why should we think our assumptions about the time reversal operator hold fixed in the context of new physical theories? What mechanism is in place in an account like Albert's, Horwich's, or Malament's to allow us to revise our conception of time reversal given enough empirical evidence? One advantage of the theory-relative approach is that it makes the question of what the time reversal operator looks like an empirical one:

---

[8]Take, for instance, the disagreement over the particle and field ontologies of quantum field theory.

since our theories are open to revision, the discrete symmetries of these theories may be too, and so our conception of what operator represents time reversal in a particular physical theory may evolve as our theory does. Thus, those swayed by empiricist or positivist considerations may think that the *a priori* character of time reversal as characterized by intuitive accounts gives us a good reason to abandon such accounts in favor of their competitors, the theory-relative accounts.

The upshot of these criticisms is simply that even the best intuitive accounts are only as good as our intuitions, and there are unfortunately many cases in contemporary physics where our intuitions should only guide us so far. So while intuitive accounts may work well in cases like classical mechanics or electromagnetism, such approaches are not suitably general to deliver the best results in cases like quantum mechanics or quantum field theory. For this reason, then, theory-relative accounts may seem the superior alternative, especially for the empiricist philosopher of physics.

However, theory-relative accounts have problems of their own. First, as pointed out by Arntzenius and Greaves, physical theories have many symmetries, and if one is willing to devote the time and effort, one can determine a class of symmetries under which any particular physical theory is invariant. However, after doing all of this work, Arntzenius and Greaves ask, why do we have any reason to call one of these operations "time reversal" in particular? If "time reversal" simply becomes synonymous with "that symmetry under which our theory is invariant", it seems like all theories are tautologously time reversal invariant and questions about time reversal are no longer of any physical interest. But let us grant that the proponent of the theory-relative account wants to provide a better definition and restricts his claims about the time reversal operator to one theory, as Roberts does with quantum mechanics. Even then, one could argue, it seems that we have no good reason to think that quantum mechanics is time reversal invariant. Why should it be? If the theory-relative account advocate accuses the intuitive account advocate of extending her intuitions of how

quantities transform under time reversal beyond their reasonable bounds, the intuitive account advocate can just as easily accuse the theory-relative account advocate of extending his intuitions of which theories are time reversal invariant beyond their reasonable bounds. As Roberts himself points out, appealing to the time reversal invariance of classical mechanics does not give anyone a good reason to presume the time reversal invariance of quantum mechanics, so what reason do we have to think quantum mechanics or any other physical theory is time reversal invariant without some basic notion of what the time reversal operator is to work from?

But the problems for theory-relative accounts run deeper, for in some cases such as those I have considered here, the theory in question underdetermines its time reversal operator. Roberts's account, for instance, provides us with two time reversal operators while Arntzenius and Greaves's textbook account seems like it *should* allow for an operator that acts as Arntzenius and Greaves's "Feynman" account suggests. This underdetermination would not be so problematic if, as with the "Feynman" and Malament accounts, according to the AG project, all such disagreements can be directly attributed to specific disagreements about fundamental ontology, but at least in Roberts's classical mechanics case there seems to be a deeper difference between the two time reversal operators; specifically, it seems that any operator that flips the position and time of a particle across an axis while leaving its velocity invariant should be called a spacetime reversal operator instead of a time reversal operator, but Roberts is content to call his T_2 a time reversal operator nonetheless. One may thus have legitimate worries that the procedure utilized in the creation of theory-relative accounts of time reversal will not yield true time reversal operators but rather a class of similar-looking symmetries, some of which may be time reversal operators and some of which may not be.

3.4 Prospects for a third way

If both the intuitive and theory-relative accounts run into the kinds of problems I have suggested, one may be tempted to look for a third way to answer questions about time reversal. Is there some way to avoid assuming an answer to the question "what does a time reversal operator look like?" while still providing an adequate answer to the question "which theories are time reversal invariant?" and vice-versa? I believe there is. Here is the basic procedure I suggest we rely on to tell us what the time reversal operator looks like in the context of a particular set of physical theories, leaving out some of the more complicated details: take the set TR_P to be composed of the physical symmetries of some set of physical theories P. The elements of TR_P must possess the following properties:

1. Each symmetry in TR_P must have some basic intuitive properties, such as being a discrete symmetry of a physical theory or being an involution.

2. Every theory in P must be invariant under a symmetry consistent with every physical symmetry in TR_P.[9]

If TR_P is the null set, then the theories in P are not invariant under time reversal. If TR_P is non-empty, then we can rely on consistency to help us identify elements of TR_P with one another. For instance, the symmetry that takes a point (x, y) to $(x, -y)$ and the symmetry that takes a point (x, y, z) to $(x, -y, z)$ may be identified with one another while the symmetry that takes a point (x, y) to $(x, -y)$ and the symmetry that takes a point (x, y, z) to $(-x, -y, z)$ may not be. If we can use the consistency criterion to identify every element of TR_P with every other element of TR_P, then the elements of TR_P are physical symmetries representing the same overarching time reversal symmetry in different physical theories. If there is no such way to unify all

---

[9]Clearly, much more must be said about what it is for a theory to be invariant under a certain symmetry, and I will discuss this issue in more detail in the next chapter.

of the elements of TR_P, then the set TR_P picks out a number of different overarching time reversal symmetries, no one of which is *the* time reversal symmetry. Either way, so long as TR_P is non-empty, we can say that the set of laws P is invariant under time reversal.

The intuition that my "third way" approach relies on is that time reversal operators in one theory can act as a constraint on the proper time reversal operator of another theory. We can use this intuition as a constraint on what time reversal should generally look like, but we must also rely on some minimal intuitions about what it means to be time reversal (as opposed to, say, parity reversal) as well. Because my account relies on such intuitions, my account does not avoid all of the problems that the limited nature of our intuitions poses for an account of time reversal, but I believe it minimizes them since the intuitions appealed to by my account ascribe features to the time reversal operator that the intuitive accounts I have discussed (as well as the theory-relative accounts, for that matter) agree upon. Likewise, the assumption that physical symmetries ought to act the same way on the same properties if they are to be called overarching-symmetric counterparts of one another seems well-motivated given my arguments in the previous chapter. Thus, the assumptions of my account are less questionable than our intuitions about the time reversal invariance of quantum field theory or our intuitions about how the time reversal operator should operate on a quark's strangeness. What is more, my account allows the question of what the time reversal operator looks like as well as which theories are invariant under time reversal to both be empirical questions that are typically discovered together, which is a result that no other account of time reversal can boast. There is something for everyone to love (and hate) about this third way; intuitive account proponents will like that their intuitions come into play when comparing the constraints placed on the time reversal operator across numerous physical theories, and theory-relative account proponents will like that we determine the time reversal operator by paying attention

to which symmetries a particular physical theory is invariant under and using these symmetries as constraints in our analysis.

Most importantly, my approach need not assume anything about which particular theories are time reversal invariant, how the time reversal operator acts on each property, what the correct ontology of a theory is, or whether or not time is handed in order to derive its results. My hope, then, is that even though this holistic and mathematically involved project will be difficult and time-consuming, it may be able to deliver non-question-begging answers to the kinds of questions we'd like an account of time reversal to answer.

CHAPTER IV

Methodology

In the first chapter I briefly proposed a new approach to the philosophical study of time reversal that requires one to calculate the discrete symmetries under which a number of theories are invariant before ruling whether these theories are time reversal invariant. Given that our best physical theories rely on differential equations, I must provide some way of determining all of the discrete symmetries under which solutions to a set of differential equations are invariant for my theory to be feasible.[1] Luckily for me, a contemporary mathematician has done most of the hard work already. In this chapter I present a brief overview of the work of *Hydon* (1998a,b, 2000a,b) and his methods for determining the discrete symmetries of ordinary and partial differential equations. At the end of the chapter, I discuss a few examples that Hydon presents in his book and discuss how my account of time reversal derives a proposal for a time reversal operator from Hydon's work. These examples are worked through in detail in preparation for my analysis in chapter 5, where my work proceeds using the same methodology but with many more examples and much more quickly. I begin, however, with a short discussion of some minimal preliminary desiderata for a time reversal operator.

---

[1]See the discussion of theories and differential equations in chapter 2 for more information on my notion of "physical theory" and the role of differential equations in my account.

4.1 A few desiderata for a time reversal operator

Though I do not agree with the general approach of intuitive accounts of time reversal for reasons discussed in the previous chapter, the basic point that one must start with some minimal notion of time reversal in order to answer either of our questions of interest is certainly true. In what follows I present some minimal constraints that must be satisfied in order for a time reversal operator to be a *time reversal* operator (as opposed to, say, a parity reversal operator, or a spacetime reversal operator). First of all, a time reversal operator should be a discrete point symmetry of time reversal invariant systems, meaning that it should be function from solutions (represented as sets of points on some variable space) to solutions which, unlike continuous symmetries like translations and rotations, shouldn't depend on some variable parameter. This follows intuitively from the fact that we never talk about time reversing some state of affairs to some extent or by some value in the way that we talk about, say, translating some figure to the left three meters or rotating some figure by an angle of ninety degrees. Time reversal also ought to be an involution, meaning that two time reversals leave the system where it started. Stated formally, this constraint requires that $T(T(\Psi)) = \Psi$ for every solution Ψ. This is again an obvious constraint on a time reversal operator since two applications of the time reversal operator should leave us with the same state we started with.

But requiring that T be an involution is not enough, for there are other physical symmetries that satisfy this condition as well. The most notable of these, parity reversal, I will represent by the operator P. The parity reversal operator is essentially the spatial analogue of time reversal. To distinguish the two, then, I simply impose the following restrictions: a physical state Ψ must inhabit the same spatial location as its time-reversed counterpart and the same "temporal location" as its parity-reversed counterpart[2]. That is, following my conventions from chapter 3, $T(\mathbf{x_n}) = \mathbf{x_n}$ and

---

[2]Clearly this constraint only applies when states Ψ are associated with particular regions in space

70

$P(t) = t$. Any candidate operator that satisfies both conditions in a case where Ψ is a function of time and position only will be called the "trivial" or "identity" operator I since it maps each state to itself. What is more, any candidate for T or P that does not satisfy either of these criteria may still be what I will call a "spacetime reversal operator", which I'll represent as A, which satisfies the condition $A(\mathbf{x_n}, t) = (u_1(\mathbf{x_n}, t), u_2(t, \mathbf{x_n}))$ for some function u_1 which satisfies $u_1(u_1(\mathbf{x_n}, t), u_2(t, \mathbf{x_n})) = x$ and for some function u_2 which satisfies $u_2(u_2(t, \mathbf{x_n}), u_1(\mathbf{x_n}, t)) = t$ (since we require that A must still be an involution). Finally, candidate operators which do not satisfy the conditions for P, T, I, or A will simply be called "general discrete symmetries", which I will represent by D. In looking for time reversal operators, then, I restrict my attention to only those operators that satisfy the condition that $T(\mathbf{x_n}) = \mathbf{x_n}$. This is not to say that parity reversal and spacetime reversal are not philosophically interesting or deeply connected with the project of constructing a proper account of T; rather, it's just that A and P are different operators from T and are thus not the subject of this particular inquiry. I do believe that, should no candidate T operators be forthcoming, candidate A operators may provide the closest thing to a time reversal operator we may find, but since the condition for T fits better with our intuitions of what separates time reversal from other involutions, we should look for T-candidates before we go searching for A-candidates to play the role of the time reversal operator in our theory.

This is all the analysis of what we're looking for in a time reversal operator I will provide. The only other constraint, consistency, will be applied only after we have determined which of our physical theories are invariant under which discrete symmetries, so I will turn next to how to determine which discrete symmetries our theories are invariant under.

---

and time. For a theory whose physical states are not functions of space and/or time, the conditions imposed by this paragraph on T, P, and A can be ignored; what will guide us to distinguishing time reversal from parity reversal in such situations will be how well certain discrete symmetries line up with parity reversal operators and time reversal operators in similar theories.

4.2 Hydon's method for ODEs

We now know some basics about the kind of symmetries we're looking for when we look for time reversal symmetries, but in order to determine whether a given theory is time reversal invariant, we first need to know what discrete symmetries the theory is invariant under. Hydon provides a clear and (relatively) simple method for calculating all of the discrete symmetries of a given differential equation. In this section I cover Hydon's approach to calculating the discrete symmetries of ordinary differential equations (ODEs), and in the next section I discuss how Hydon's approach extends to partial differential equations (PDEs).

We begin with the fact that every ODE of the form:[3]

$$x^{(n)} = \omega(t, x, x', ..., x^{(n-1)}) \tag{4.1}$$

where $x^{(n)}$ is the n-th derivative of x with respect to t and $n \geq 2$, has a finite-dimensional Lie algebra L of continuous point symmetry generators. The basis for these generators can be written in the following form:[4]

$$X_i = \xi_i(t, x)\partial t + \eta_i(t, x)\partial x \tag{4.2}$$

Any symmetry generator in L can be exponentiated to yield a one-parameter Lie group of symmetries of the given differential equation that are continuously connected to the identity component of the symmetry group. Hydon proves that one

---

[3]I have slightly altered these examples from Hydon's conventions, replacing the independent variable x with t and the dependent variable u with x to make it clear where the time coordinate enters into these equations. It is assumed here that t is the independent variable and x is the dependent variable.

[4]The actual generators may be much longer, including terms of $\eta_{i(k)}\partial x^{(k)}$, but I omit those higher-order terms for the moment for the sake of simplicity.

can derive all of the discrete point symmetries of an ODE once one has determined the elements in the basis of L, which requires that one solve for $\xi_i(t, x)$ and $\eta_i(t, x)$. That is, once we learn the basis of the generators for the symmetry groups of the ODE continuously connected to the identity, we can determine all of the discrete symmetries of the ODE in question. So my analysis here will proceed in two steps: I'll begin by quickly explaining how Hydon determines the Lie group of symmetries of a differential equation, and then I'll explain Hydon's novel contribution to the literature where he explains how to extend the analysis of Lie symmetries in order to determine all of the discrete symmetries of a differential equation.

In order to determine the functions $\xi_i(t, x)$ and $\eta_i(t, x)$, Hydon returns to the original notion of a symmetry operator, which is an operator Γ such that:

$$\Gamma : (t, x, x', ..., x^{(n-1)}) \rightarrow (\hat{t}, \hat{x}, \hat{x}', ..., \hat{x}^{(n-1)}) \tag{4.3}$$

where the coordinates $(\hat{t}, \hat{x}, \hat{x}', ..., \hat{x}^{(n-1)})$ must satisfy the equation:

$$\hat{x}^{(n)} = \omega(\hat{t}, \hat{x}, \hat{x}', ..., \hat{x}^{(n-1)}) \tag{4.4}$$

since symmetries are maps from solutions of an ODE to solutions of the same ODE. This equation is called the *symmetry condition* of the ODE, and while alone it is not enough to derive the basis for L, it's a start.

Consider now a symmetry that leaves every solution to the ODE infinitesimally changed. For such a symmetry, Hydon claims, the variables $(\hat{t}, \hat{x}, \hat{x}', ..., \hat{x}^{(n-1)})$ can be written out in terms of $(t, x, x', ..., x^{(n-1)})$ as follows:

$$\hat{t} = t + \epsilon\xi + O(\epsilon^2) \tag{4.5}$$

$$\hat{x} = x + \epsilon\eta + O(\epsilon^2) \tag{4.6}$$

$$\hat{x}_{(k)} = x_{(k)} + \epsilon\eta_{(k)} + O(\epsilon^2) \tag{4.7}$$

where ϵ is close to zero and the k in $\eta_{(k)}$ is merely an index. Plugging the terms of order ϵ into the symmetry condition yields the following, called the *linearized symmetry condition*, which will allow us to determine a basis for the generators of the symmetry group continuously connected to the identity component:

$$\eta_{(k)} = \xi\omega_t + \eta\omega_x + \eta_{(1)}\omega_{x'} + ... + \eta_{(k-1)}\omega_{x^{(k-1)}} \tag{4.8}$$

Identities involving the $\eta_{(k)}$ terms may be substituted into this equation depending on the ODE in question (examples of these sorts of conditions will be discussed later on in this paper when I consider examples of Hydon's method in action), and because all of the η and ξ terms are independent of higher-order powers of x, this linearized symmetry condition can typically be broken into a system of determining differential equations that allows us to solve for ξ and η and thus determine a basis of L.

One quick thing to note here: once the basis has been derived, one can then determine the commutation relations among elements of the basis. That is, one can derive the constants c_{ij}^k in the following equation:

$$[X_i, X_j] = c_{ij}^k X_k \tag{4.9}$$

The c_{ij}^k constants are of interest, as we'll see shortly, because every symmetry transfor-

mation Γ induces an automorphism of L which preserves these commutation relations; that is, the commutation relations among the elements of the basis of L are invariant under the symmetries of the ODE.

Up to this point my analysis, following Hydon's, has followed the standard discussion of how to derive the basis of L for a given ODE as found in textbooks like that of *Olver* (1993). This basis allows us to determine all of the continuous symmetries of a differential equation, but it does not straightforwardly tell us the discrete symmetries of that differential equation since the discrete symmetries form a subgroup of the symmetry group that is not continuously connected to the identity component and so for which the linearized symmetry condition does not hold. The rest of this section will cover Hydon's novel contribution to the literature by outlining his method for determining all of the discrete symmetries under which a particular ODE is invariant. Hydon begins his novel analysis by proving a lemma that shows that the adjoint action of any symmetry in the full symmetry group of the differential equation (which includes both continuous and discrete symmetries) induces an automorphism on L. The adjoint action of the symmetry S, for instance, on the elements of the basis of L, X_i, can be written as $\hat{X}_i = S^{-1} X_i S$. Hydon shows that the new generators $\hat{X}_i$, the result of the adjoint action of S on X_i, form a basis for L, and so the original basis elements X_i can be written in terms of the new basis elements $\hat{X}_i$ as follows:

$$X_i = b_i^l \hat{X}_l \tag{4.10}$$

For ease of calculation it is useful to take these coefficients as the elements of the non-singular matrix $B = (b_i^l)$. Each element $\hat{X}_i$ can itself be written as follows:

$$\hat{X}_i = \xi_i(\hat{t}, \hat{x})\partial\hat{t} + \eta_i(\hat{t}, \hat{x})\partial\hat{x} \tag{4.11}$$

Assuming that there are G basis elements, the first of the above conditions yields $2G$ determining equations of the following form:

$$X_i\hat{t} = b_i^l \xi_l(\hat{t}, \hat{x}) \tag{4.12}$$

$$X_i\hat{x} = b_i^l \eta_l(\hat{t}, \hat{x}) \tag{4.13}$$

These $2G$ equations, along with the original symmetry condition, are enough for us to determine all of the symmetries represented by Γ above, assuming that we already know B. So, to quickly recap: elements of the full symmetry group (including discrete symmetries) induce a transformation on the X basis, and so once we know these elements of L, we can solve a series of determining equations that provide us with a mapping from our original variables $(t, x, ...)$ to the symmetry-transformed counterparts of these variables $(\hat{t}, \hat{x}, ...)$ for all of symmetries of the differential equation we're interested in.

One technical mystery of how this procedure works yet remains: how do we determine the elements of the matrix B? Hydon shows that, taking the fact that $\Gamma : X_i \rightarrow b_i^l \hat{X}_l$ into account along with the fact that the commutation relations remain unchanged by Γ, we can obtain the following system of nonlinear equations that determines the elements of B given the constants c_{ij}^k that appear in the commutation relations:

$$c_{lm}^n b_i^l b_j^m = c_{ij}^k b_k^n \tag{4.14}$$

76

Note that, in order avoid double-counting equations, we need only consider the equations where $1 \leq i < j \leq G$.

But our work is not quite done. Hydon's method successfully gives us all of the elements of the symmetry group by helping us determine those symmetries whose adjoint action links the generators X_i and $\hat{X}_i$; but we are looking for just the discrete symmetries, not all of the symmetries of these differential equations, so Hydon ends his method by requiring that we factor out the continuous symmetries in a series of explicit steps too involved for me to discuss here at length.[5]

So, in short, the following is Hydon's method for finding all of the discrete symmetries of an ODE:

1. Determine X_i, the basis of L, using the linearized symmetry condition.

2. Find the commutation relations among X_i in order to determine the constants c_{ij}^k.

3. Use the constants c_{ij}^k to determine the elements of the matrix B.

4. Use the elements of the matrix B, along with the symmetry condition, to solve for $(\hat{t}, \hat{x})$.

5. Factor out the continuous symmetries, leaving only the discrete symmetries of the system.

Clearly I have glossed over many of the details of Hydon's account in an attempt to streamline it and not get too bogged down in the details; however, it is probably already apparent just how difficult this procedure will be to execute for most ordinary differential equations, especially for ones that are highly symmetric. In order to cut through some of this difficulty, then, I will turn, later in this chapter, to the application of my account of time reversal to some results Hydon has already worked through in order to see just what verdicts my account delivers in some toy cases.

---

[5]Interested readers are referred to *Hydon* (1998a).

4.3 Hydon's method for PDEs

The five-step methodology for determining all of the discrete symmetries of ODEs outlined in the previous section carries over, generally and for the most part, to PDEs. As with ODEs, the first step in this procedure involves calculating the elements X_i from a linearized symmetry condition; however, the linearized symmetry condition for PDEs is fundamentally different from the linearized symmetry condition for ODEs. Aside from this difference in the first step of the procedure, all of Hydon's methods for calculating the discrete symmetries of ODEs carry over to the calculation of discrete symmetries of PDEs; that is, once one has succeeded in calculating the elements X_i, it doesn't matter whether one obtained the elements X_i from an ODE or a PDE: Hydon's procedure follows just the same. Thus, in this section, I will only focus on a quick derivation of the linearized symmetry condition for PDEs.

Start out with the simplifying assumption that one is working with a PDE Q that has one dependent variable, u, and two independent variables, x and t[6]. The elements of L, the lie algebra of point symmetry generators continuously connected to the identity element for Q, will have the following form:

$$X_i = \xi(x,t,u)\partial_x + \tau(x,t,u)\partial_t + \eta(x,t,u)\partial_u \tag{4.15}$$

Again, as with ODEs, we have a series of unknowns (ξ, τ, η) for whose solution we must find the determining equations. Again, we rely on infinitesimal point symmetry transformations of the form:

---

[6]The generalization of this analysis to PDEs with more dependent or independent variables follows straightforwardly.

$$\hat{x} = x + \epsilon \xi(x, t, u) + O(\epsilon^2) \tag{4.16}$$

$$\hat{t} = t + \epsilon \tau(x, t, u) + O(\epsilon^2) \tag{4.17}$$

$$\hat{u} = u + \epsilon \eta(x, t, u) + O(\epsilon^2) \tag{4.18}$$

$$\hat{u}_J = u_J + \epsilon \eta_J(x, t, u) + O(\epsilon^2) \tag{4.19}$$

where u_J is the partial derivative of u with respect to some combination of x and t labeled J. I also follow my previous notation here in using η_J to represent prolongations of η, not derivatives or powers of η with respect to these variables. These prolongations can be calculated on a case-by-case basis by applying total derivatives of the form:

$$D_x = \partial_x + u_x \partial_u + u_{xx} \partial_{u_x} + u_{xt} \partial_{u_t} + ... \tag{4.20}$$

$$D_t = \partial_t + u_t \partial_u + u_{xt} \partial_{u_x} + u_{tt} \partial_{u_t} + ... \tag{4.21}$$

Many of these prolongations are also calculated explicitly on page 141 of Hydon's textbook, and so I will not explicitly calculate any of the prolongations commonly utilized in symmetry derivations here. The linearized symmetry condition requires that $\hat{x}$, $\hat{t}$, and $\hat{u}$ satisfy Q just as well as x, t, and u do. All terms of order $O(\epsilon^2)$ are ignored, just as in the ODE case, and one obtains a symmetry condition from the terms of order ϵ.

Consider the following example of Burgers' equation:[7] $u_t + uu_x = u_{xx}$

---

[7]This example is taken from *Hydon* (2000a), p. 144

$$\hat{u}_t + \hat{u}\hat{u}_x = \hat{u}_{xx} \tag{4.22}$$

$$u_t + \epsilon\eta_t + (u + \epsilon\eta)(u_x + \epsilon\eta_x) = u_x x + \epsilon\eta_{xx} \tag{4.23}$$

The terms of order ϵ can be isolated from the other terms in this equation which are straightforwardly seen to be equivalent via Burger's equation. Thus, the linearized symmetry condition, found by equating terms of order ϵ, is simply the following:

$$\eta_t + u_x\eta + u\eta_x = \eta_{xx} \tag{4.24}$$

These prolongations of η can be written in terms of η, τ, and ξ, providing us with a system of generating equations that allow us to determine the elements of L. As previously stated, once the elements X_i have been determined, the discrete symmetries can be calculated just as they were in the previous section.

4.4 Hydon's solutions

Carrying out the procedure outlined in the previous sections is, for many differential equations (especially highly symmetric ones), tough work. In order to evaluate the promise of the method I've proposed, one needs examples of the results it can deliver in many interesting cases. Normally, it would take quite a lot of effort to conduct such an analysis, but Hydon has done us a favor. In chapter 11 of his book, *Hydon* (2000a) provides three differential equations all of whose discrete symmetries he explicitly calculates. In this section I will discuss these three examples and, using the criteria I proposed at the beginning of the chapter, determine which of these discrete symmetries seem to be the best candidates for the time reversal operator if we

take all three of these equations to be true or to represent the best physical theories we have available to us. Of course, these three equations, though interesting, do not necessarily fulfill either of these two criteria (we're dealing with toy cases here, after all), but at the very least these equations will reveal just how my proposal can use the results of Hydon's procedure to determine the best or most useful form of the time reversal operator.[8]

4.4.1 Example 11.4

Hydon's first full example in Chapter 11 is the following ODE:[9]

$$x''' = \frac{x''^2}{t} - \frac{x''}{x'}$$ (4.25)

where the primed terms are derivatives with respect to time. Hydon goes through the steps outlined in the previous section and arrives at the following discrete symmetries:

$$S_1 \Psi(x, t) = \Psi(x, t)$$ (4.26)

$$S_2 \Psi(x, t) = \Psi(x, -t)$$ (4.27)

4.4.2 The Chazy equation

Hydon's next example in Chapter 11 is the Chazy equation, an ordinary differential equation primarily of interest to mathematicians because it is the simplest example of an ordinary differential equation with solutions that have a moveable natural boundary. It can be written as follows:

---

[8]It is worth noting that all of the examples I consider here are two-dimensional (one spatial dimension, one temporal) for the sake of simplicity.

[9]For the purpose of these examples I will interpret the variables x, across all equations considered here, as representing the position of a particular particle in space at some time t.

$$x''' = 2xx'' - 3x'^2 + \lambda(6x' - x^2)^2 \tag{4.28}$$

Again, Hydon goes through the steps outlined in the previous section and arrives at the following discrete symmetries:

$$S_1 \Psi(x,t) = \Psi(x,t) \tag{4.29}$$

$$S_3 \Psi(x,t) = \Psi(-x,-t) \tag{4.30}$$

$$S_4 \Psi(x,t) = \Psi(xt^2 + 6t, -\frac{1}{t}) \tag{4.31}$$

$$S_5 \Psi(x,t) = \Psi(-(xt^2 + 6t), \frac{1}{t}) \tag{4.32}$$

4.4.3 The Harry-Dym equation

Hydon's final example in Chapter 11 is the Harry-Dym equation, a nonlinear wave equation primarily of interest to mathematicians and physicists for its connection to the Korteweg-de Vries equation but not, to my knowledge, due to any independent physical significance. The Harry-Dym equation is a partial differential equation that can be written as follows:

$$u_t = u^3 u_{xxx} \tag{4.33}$$

In this case, u represents some function of t and x, and u_t represents the partial derivative of u with respect to t while u_{xxx} represents the third partial derivative of u with respect to x. Again, Hydon goes through the steps outlined in the previous section and arrives at the following discrete symmetries:

$$S_6\Psi(x,t,u) = \Psi(x,t,u) \tag{4.34}$$

$$S_7\Psi(x,t,u) = \Psi(-x,t,-u) \tag{4.35}$$

$$S_8\Psi(x,t,u) = \Psi(x,-t,-u) \tag{4.36}$$

$$S_9\Psi(x,t,u) = \Psi(-x,-t,u) \tag{4.37}$$

$$S_{10}\Psi(x,t,u) = \Psi(-\frac{1}{x},t,\frac{u}{x^2}) \tag{4.38}$$

$$S_{11}\Psi(x,t,u) = \Psi(-\frac{1}{x},-t,-\frac{u}{x^2}) \tag{4.39}$$

$$S_{12}\Psi(x,t,u) = \Psi(\frac{1}{x},t,-\frac{u}{x^2}) \tag{4.40}$$

$$S_{13}\Psi(x,t,u) = \Psi(\frac{1}{x},-t,\frac{u}{x^2}) \tag{4.41}$$

4.4.4 Time reversal in Hydon's examples

The preceding sections show that Hydon has provided us with 13 candidates for the time reversal operator if we think that most (if not all) of the three equations examined are to be time reversal invariant. In order to determine which of these 13 candidates actually provides a good candidate for T, remember the following minimal constraints I placed on a time reversal operator back in section 4.1:[10]

$$T(T\Psi(x,t)) = \Psi(x,t) \tag{4.42}$$

$$T(x) = x \tag{4.43}$$

The first of these constraints requires that the time reversal operator be an involution while the second separates time reversal from parity reversal and spacetime reflection. However, even before applying these constraints we can reject S_1 and S_6 as candidates

---

[10]The second of these is not quite the same as the one discussed in section 4.1 but it is its analogue for cases in which the states in question are functions of a single spatial parameter.

for T because S_1 and S_6 are simply identity operators: they map solutions of the differential equations considered to themselves, so these symmetries are trivial and cannot be reasonable candidates for T.

The second of the above minimal constraints is powerful enough to eliminate a number of the above T-candidates. For instance, S_3, S_7, and S_9 all involve flipping the sign of x while S_4 and S_5 and S_{10}, S_{11}, S_{12}, and S_{13} involve more complicated changes to x. As such, all of these candidates are unfit to serve as T. This leaves only S_2 and S_8 as candidates for T. Both of these candidates are also involutions, meaning that they satisfy the first criterion listed above as well.

The final constraint on T involves the fact that we are looking for a time reversal operator that is consistent across theories. So, if the same quantities appear in multiple differential equations, then the time reversal operator, applied to the solutions of these differential equations, should change the quantities in a consistent way. In this particular case, where we consider the following T-candidates:

$$S_2\Psi(x,t) = \Psi(x,-t) \tag{4.44}$$

$$S_8\Psi(x,t,u) = \Psi(x,-t,-u) \tag{4.45}$$

it is easy to see that the symmetries represented by S_2 and S_8 clearly operate on their common variables x and t in the same way: both take x to x and take t to $-t$. In the language introduced in chapter 2, we could say that these two symmetries, though different physical symmetries, can be treated as instances of the same overarching symmetry. We are not always so lucky, however. Imagine that, instead of S_6, we had found that S_6' listed below was a symmetry of one of our differential equations:

$$S_6' \Psi(x, t, u) = \Psi(x, t, -u) \qquad (4.46)$$

We could not have ruled this symmetry out as a candidate for T previously, but my new criterion allows us to. The symmetry S_6' differs from S_2 because t does not change sign under S_6', so S_8 and S_2 are consistent while S_6' cannot be consistently labeled as the same operator, meaning that one should take S_8 to be the time reversal operator given these equations and my analysis since it is consistent with S_2 but contains more information than S_2. Therefore, my criterion has allowed me to pick out a unique time reversal operator from this set of differential equations, and even though this criterion was of little help in this specific example, it is clear how, in cases where symmetries like S_6' are under consideration, my criterion allows us to determine which symmetries are good universal candidates for T and which ones are not.

The fact that my proposal for the correct way to analyze the time reversal operator has provided a unique best candidate for T certainly tells in its favor, but note that there is an interesting consequence to my analysis: if the three differential equations described above are taken to be the fundamental laws of nature in our toy universe, then these laws are not jointly time reversal invariant. This follows from the fact that S_2 and S_8 are not discrete symmetries of the Chazy equation. Therefore, we have no consistent way of making all of these fundamental laws time reversal invariant. There is perhaps little of interest in finding that a toy universe has fundamental laws that are not time reversal invariant, but if the fundamental laws of our own universe were not time reversal invariant, this would certainly be an interesting result that flies in the face of much conventional wisdom about fundamental physics.

In the face of such a problem, one may be interested in examining potential candidates for a spacetime reversal operator A among the options considered; after

all, perhaps, in the same way that the laws of special relativity show that there is an objective, frame-independent notion of a spacetime interval but not of a temporal interval, our fundamental physical theories may reveal that there is an objective notion of spacetime reversal under which all fundamental physical theories are invariant but no such notion of time reversal. However, in the example considered above, the prospects for such a common A under which all three equations are invariant are extremely poor. The equation in Hydon's example 11.4 is invariant only under S_1 and S_2, and I have already shown that S_1 is the identity operator I while S_2 is a simplified version of the time reversal operator S_6. Thus, even if there were a consistent spacetime reversal operator between the Chazy and Harry-Dym equations (S_9 seems like a good candidate), it would still not establish a candidate for A under which all of these "laws" are invariant. Still, it is useful to consider such a line of inquiry since, though it yields no fruit when applied to these three equations, it may do so for other sets of differential equations.

Before leaving this example, however, there is one way to potentially save the time reversibiltiy of the fundamental laws of this toy universe that I should mention: one could argue that my original assumption that the variables x and t in the three equations above represent the same physical property was false. Of course, this being a toy universe, I can simply stipulate that x and t are pick out the same physical features across the three equations I considered, but to establish such a claim with respect to actual physical theories may be harder. Not all theories wear their interpretation functions on their sleeves. So perhaps when one is dealing with actual examples of physical laws, one can reject the final step of the procedure I implemented above by arguing that the variables x and t in different laws pick out different physical features in the world. Clearly this maneuver should not be implemented blindly, but I note it because it is a response to my analysis that needs to be kept in mind as I move from toy cases to real cases lest one mistake a similarity in the variables conventionally

assigned to two physical properties for an identity between the two properties in question.

In the next chapter, I'll provide a comprehensive list of differential equations all of whose discrete symmetries have been determined to date. I'll then examine sets of these differential equations to determine what time reversal would look like under my account in a world governed by these differential equations. This analysis, I'll argue. gives us a defeasible reason to believe that the fundamental laws of physics are not time reversal invariant on my account.

CHAPTER V

An Application of the Third Way to a Series of Differential Equations

The previous chapters introduced what I have called the "third way", a new approach to time reversal that requires one to examine the differential equations utilized by a number of different physical theories and the discrete symmetries under which these differential equations are invariant in order to determine both what form the a time reversal transformation should take and what physical theories are invariant under that transformation. My methodology utilizes work by *Hydon* (1998a, 1999, 2000a,b, 2001) and suggests implementing his techniques in order to determine all of the discrete symmetries under which a given differential equation is invariant. While implementing Hydon's methodology may be difficult in some cases and may not always be useful,[1] Hydon and a number of others (namely *Levi and Rodríguez* (2004), *Silberberg* (2005), and *Hamad et al.* (2011)) have produced all of the discrete symmetries for number of differential equations. In this chapter, I will draw on the work of these authors, as well as my own original work, to analyze the discrete symmetries of seventeen differential equations and, more importantly, the involutions

---

[1]For instance, there may be cases where a differential equation is not invariant under any continuous Lie symmetry transformation but is invariant under some discrete symmetry transformation, and Hydon's method would not provide us with any discrete symmetries in such a case. Such cases are, however, quite rare, and as long as a differential equation is invariant under *some* continuous Lie symmetry transformation, Hydon's method will deliver all of the discrete symmetries under which that differential equation is invariant.

under which these differential equations are invariant. I will take this catalogue of involutions and, in the third section of this chapter, discuss what my "third way", when applied to this catalogue of involutions, tells us about the time reversal operator in worlds governed solely by select subsets of the equations I will analyze. Finally, I will conclude by drawing comparisons between my analysis and the analysis suggested by both "intuitive" and "theory-relative" time reversal theorists. Those who are not interested in a more thorough analysis of all of the discrete symmetries of the differential equations I discuss in this paper are invited to skip to the third section and refer to Tables 1 and 2 as a summary of the second section's results.

5.1 The Equations

For each of the following equations, I will use the shorthand u_x to stand for $\frac{\partial u}{\partial x}$, u_{xx} as shorthand for $\frac{\partial^2 u}{\partial x^2}$, and so on. In cases with one dependent variable and one independent variable, I will call the dependent variable x and the independent variable t, while in cases where there are one dependent and two independent variables, x and t will refer to the independent variables while u will refer to the dependent variable. I ignore this convention in several notable cases, such as the Gas equation, which is unique in that it has many more variables than the other equations I analyze here, and in several cases involving change of coordinates such as Equation Q and the Inverse Square Law. In these cases, I try to keep my notation as close to the original author's as possible. My notation elsewhere is not shared by all of the authors whose work I cite, but it makes it easier to see similarities across the discrete symmetries I analyze.

5.1.1 Black-Scholes Equation

The Black-Scholes Equation is:

$$u_t + \frac{1}{2}A^2 x^2 u_{xx} + Bxu_x - Cu = 0 \tag{5.1}$$

For some constants A, B, and C. It has the following non-trivial discrete symmetries, according to *Silberberg* (2005):

$$T_{A1}(x, t, u) = \left(e^{2Dt - log(x)}, t, u\right) \tag{5.2}$$

$$T_{A2}(x, t, u) = \left(e^{A^{-2}D - A^{-4}Dt^{-1} - A^{-2}t^{-1}log(x)}, -A^{-4}t^{-1}, A\sqrt{|t|}e^{-\frac{1}{2A^2t}[(log(x) - Dt)^2 + 2A^2Ct^2 + \frac{2C}{A}]}u\right) \tag{5.3}$$

$$T_{A3}(x, t, u) = \left(e^{-A^{-2}D - A^{-4}Dt^{-1} + A^{-2}t^{-1}log(x)}, -A^{-4}t^{-1}, A\sqrt{|t|}e^{-\frac{1}{2A^2t}[(log(x) - Dt)^2 + 2A^2Ct^2 + \frac{2C}{A}]}u\right) \tag{5.4}$$

Where $D = B - \frac{A^2}{2} \neq 0$. Of these, only T_{A1} is an involution.

5.1.2 Burgers' Equation

Burgers' Equation is:

$$u_{xx} = u_t + uu_x \tag{5.5}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (2000b):

$$T_{B1}(x, t, u) = (-x, t, -u) \tag{5.6}$$

$$T_{B2}(x, t, u) = (-ix, -t, iu) \tag{5.7}$$

$$T_{B3}(x, t, u) = (ix, t, -iu) \tag{5.8}$$

$$T_{B4}(x, t, u) = \left(\frac{x}{2t}, -\frac{1}{4t}, 2(tu - x)\right) \tag{5.9}$$

$$T_{B5}(x, t, u) = \left(-\frac{x}{2t}, -\frac{1}{4t}, -2(tu - x)\right) \tag{5.10}$$

Of these, only the first is an involution.

5.1.3 Spherical Burgers' Equation

The Spherical Burgers' Equation is:

$$u_{xx} = u_t + \frac{u}{t} + uu_x \tag{5.11}$$

It has only the following non-trivial discrete symmetry, according to *Hydon* (2000b):

$$T_C(x, t, u) = (-x, t, -u) \tag{5.12}$$

T_C is an involution.

5.1.4 Burgers' Equation with Time Dependent Flux at the Origin

The following is called Burgers' Equation with time-dependent flux at the origin by *Hamad et al.* (2011):[2]

$$u_t = u_{xx} + 2uu_x \tag{5.13}$$

It has the following non-trivial discrete symmetries:

---

[2]Note that the only difference between this equation and Burgers' equation is a factor of -2 on the uu_x term. *Hamad et al.* (2011) do not provide details on how time-dependent flux explains this extra factor, but I imagine this extra factor comes from applying the Hopf-Cole transformation $u \to \frac{2}{w}w_x$, where $w(x, t)$ is a solution to the linear heat equation $w_t = w_{xx}$, and applying the initial condition $w_t = f(t)$ for some function f at the point $x = 0$.

$$T_{D1}(x,t,u) = \left(-\frac{\sqrt{2}}{t}x, -\frac{2}{t}, -\frac{\sqrt{2}}{4}(x+2ut) \right) \tag{5.14}$$

$$T_{D2}(x,t,u) = \left(-\frac{-\sqrt{2}}{t}x, -\frac{2}{t}, -\frac{-\sqrt{2}}{4}(x+2ut) \right) \tag{5.15}$$

$$T_{D3,\lambda}(x,t,u) = (4\lambda + x, t, u) \tag{5.16}$$

$$T_{D4,\lambda}(x,t,u) = (4\lambda - x, t, -u) \tag{5.17}$$

$$T_{D5,A}(x,t,u) = \left(\frac{1}{A}x, \frac{t - \sqrt{2}A}{A^2}, Au \right) \tag{5.18}$$

$$T_{D6,A}(x,t,u) = \left(\frac{1}{A}x, \frac{t + \sqrt{2}A}{A^2}, Au \right) \tag{5.19}$$

Where λ and A are arbitrary constants. Of these, only $T_{D3,\lambda=0}$ and $T_{D4,\lambda=0}$ are involutions, and since $T_{D3,\lambda=0}$ is the identity operator, only $T_{D4,\lambda=0}$ is of interest to us.

5.1.5 The Chazy Equation

The Chazy Equation is:

$$x_{ttt} = 2xx_{tt} - 3y_t^2 + \lambda(6x_t - x^2)^2 \tag{5.20}$$

For some constant λ. It has the following non-trivial discrete symmetries, according to *Hydon* (1998a, 2000a):

$$T_{E1}(t,x) = (-t, -x) \tag{5.21}$$

$$T_{E2}(t,x) = \left(-\frac{1}{t}, t^2x + 6t \right) \tag{5.22}$$

$$T_{E3}(t,x) = \left(\frac{1}{t}, -(t^2x + 6t) \right) \tag{5.23}$$

92

All three of these symmetries are involutions.

5.1.6 Euler-Poisson-Darboux Equation

The Euler-Poisson-Darboux (EPD) Equation is:

$$u_{tt} - u_{xx} = \frac{p(p+1)}{t^2}u \tag{5.24}$$

For some integer p. It has the following non-trivial discrete symmetries, according to *Hydon* (1999):

$$T_{F1}(x,t,u) = (-x,t,u) \tag{5.25}$$

$$T_{F2}(x,t,u) = (x,-t,u) \tag{5.26}$$

$$T_{F3}(x,t,u) = (x,t,-u) \tag{5.27}$$

$$T_{F4}(x,t,u) = (-x,-t,u) \tag{5.28}$$

$$T_{F5}(x,t,u) = (x,-t,-u) \tag{5.29}$$

$$T_{F6}(x,t,u) = (-x,t,-u) \tag{5.30}$$

$$T_{F7}(x,t,u) = (-x,-t,-u) \tag{5.31}$$

$$T_{F8}(x,t,u) = \left(\frac{x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, u\right) \tag{5.32}$$

$$T_{F9}(x,t,u) = \left(\frac{-x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, u\right) \tag{5.33}$$

$$T_{F10}(x,t,u) = \left(\frac{x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, u\right) \tag{5.34}$$

$$\tag{5.35}$$

$$T_{F11}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, -u \right) \tag{5.36}$$

$$T_{F12}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, u \right) \tag{5.37}$$

$$T_{F13}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, -u \right) \tag{5.38}$$

$$T_{F14}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, -u \right) \tag{5.39}$$

$$T_{F15}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, -u \right) \tag{5.40}$$

All of these symmetries are involutions.

5.1.7 Gas Equations

The gas equations are:

$$q_s - v_y = 0 \tag{5.41}$$

$$v_s + p_y = 0 \tag{5.42}$$

$$p_s + \frac{3p}{q} v_y = 0 \tag{5.43}$$

With dependent variables v, p, and q and independent variables s and y. They have the following non-trivial discrete symmetries, according to *Hydon* (2000b):

$$T_{G1}(s, y, v, p, q, \phi) = \left(-\frac{1}{s}, y, \phi - sv, -s^3 p, -\frac{q}{s}, -\frac{\phi}{s} \right) \tag{5.44}$$

$$T_{G2}(s, y, v, p, q, \phi) = (-s, y, -v, p, q, \phi) \tag{5.45}$$

$$T_{G3}(s, y, v, p, q, \phi) = (s, -y, -v, p, q, -\phi) \tag{5.46}$$

Of these, only T_{G2} and T_{G3} are involutions.

5.1.8 The Harry-Dym Equation

The Harry-Dym Equation is:

$$u_t = u^3 u_{xxx} \tag{5.47}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (1999, 2000a,b):

$$T_{H1}(x,t,u) = (x, -t, -u) \tag{5.48}$$

$$T_{H2}(x,t,u) = (-x, t, -u) \tag{5.49}$$

$$T_{H3}(x,t,u) = (-x, -t, u) \tag{5.50}$$

$$T_{H4}(x,t,u) = \left(-\frac{1}{x}, t, \frac{u}{x^2}\right) \tag{5.51}$$

$$T_{H5}(x,t,u) = \left(\frac{1}{x}, t, \frac{-u}{x^2}\right) \tag{5.52}$$

$$T_{H6}(x,t,u) = \left(-\frac{1}{x}, -t, \frac{-u}{x^2}\right) \tag{5.53}$$

$$T_{H7}(x,t,u) = \left(\frac{1}{x}, -t, \frac{u}{x^2}\right) \tag{5.54}$$

All seven of these symmetries are involutions.

5.1.9 The Heat Equation

The heat equation is:

$$u_{xx} = u_t \tag{5.55}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (1999):

$$T_{I1,C}(x,t,u) = (-x,t,Cu) \tag{5.56}$$

$$T_{I2,C}(x,t,u) = (x,t,Cu) \tag{5.57}$$

$$T_{I3}(x,t,u) = \left(\frac{x}{2t}, -\frac{1}{4t}, \sqrt{2it}e^{\frac{x^2}{4t}}u\right) \tag{5.58}$$

$$T_{I4}(x,t,u) = \left(-\frac{x}{2t}, -\frac{1}{4t}, \sqrt{2it}e^{\frac{x^2}{4t}}u\right) \tag{5.59}$$

where C is a complex constant. Of these, $T_{I1,C=-1}$, $T_{I1,C=1}$, and $T_{I2,C=-1}$ are the only non-trivial involutions.

5.1.10 Toda Equation

The Toda Equation, a discrete differential equation, is:

$$u_{tt} = e^{u_+ - u} + e^{u - u_-} \tag{5.60}$$

where $u_\pm = u(x \pm h, t)$ and h is the lattice step which separates two adjacent values of x. It has the following non-trivial discrete symmetries, according to *Levi and Rodríguez* (2004):

$$T_{J1}(x,t,u) = (x,-t,u) \tag{5.61}$$

$$T_{J2}(x,t,u) = (-x,t,-u) \tag{5.62}$$

$$T_{J3}(x,t,u) = (-x,-t,-u) \tag{5.63}$$

All three of these symmetries are involutions.

5.1.11 The Free Particle

Laine-Pearson and Hydon (2003) provide an extremely useful taxonomy of discrete

symmetries under which differential equations with a particular group of symmetry generators are invariant. The generators provided on page 48 of *Hydon* (2000a) for the following differential equation:

$$x_{tt} = 0 \tag{5.64}$$

thus allow us to use the work of *Laine-Pearson and Hydon* (2003) to determine the discrete symmetries of the free particle governed by this equation. The discrete symmetries of the free particle is the 24-element group $S(4)$, so in the interests of space, I will only list the non-trivial involutions under which the free particle is invariant below:

$$T_{K1}(t, x) = (-t, x) \tag{5.65}$$

$$T_{K2}(t, x) = (t, -x) \tag{5.66}$$

$$T_{K3}(t, x) = (-t, -x) \tag{5.67}$$

$$T_{K4}(t, x) = (x, t) \tag{5.68}$$

$$T_{K5}(t, x) = (-x, -t) \tag{5.69}$$

$$T_{K6}(t, x) = \left(\frac{t}{x}, \frac{1}{x}\right) \tag{5.70}$$

$$T_{K7}(t, x) = \left(-\frac{t}{x}, \frac{1}{x}\right) \tag{5.71}$$

$$T_{K8}(t, x) = \left(\frac{1}{t}, \frac{x}{t}\right) \tag{5.72}$$

$$T_{K9}(t, x) = \left(\frac{1}{t}, -\frac{x}{t}\right) \tag{5.73}$$

5.1.12 Newtonian Inverse Square Law in a Single Dimension

The equation for a Newtonian particle operating under the influence of a force proportional to the inverse square of the distance between the force's source and the

particle in one dimension is:

$$x_{tt} = \frac{1}{x^2} \tag{5.74}$$

I calculated the discrete symmetries of this equation myself, using the YaLie package for Mathematica, and found the following to be the sole non-trivial involutive symmetry of the equation:[3]

$$T_{L1}(t, x) = (-t, x) \tag{5.75}$$

This may seem strange to some since there is no good parity reversal candidate (that is, no good candidate that acts on y while leaving x invariant) under which the inverse square law is invariant, but note that the particular form of the inverse square law given above is not the one we are used to using within physics. If we read x as position relative to a particular reference point, as the selection of this variable invites us to do, we recognize the fact that x can take on both negative and positive values. But inverse square laws are written in polar coordinates because what is physically significant is the magnitude of the distance between the two points, not the orientation of one object with respect to the other. So, in the one-dimensional case, we utilize the variable substitution $r = \sqrt{x^2} = |x|$. A standard parity reversal operator which, say, takes x to x and y to $-y$ is not a symmetry of the above equation because such a transformation takes y_{xx} to $-y_{xx}$, reversing the sign of the left-hand side of the equation while leaving y^2 and thus the right-hand side of the equation unchanged. If we convert our equation into the correct formulation of the inverse square law in polar coordinates, we get the following:

$$r_{tt} = \frac{1}{r^2} \tag{5.76}$$

---

[3]For documentation of this Mathematica package, please see *Díaz* (2003).

or

$$|x|_{tt} = \frac{1}{x^2} \tag{5.77}$$

which is invariant under T_{L1} and invariant under the following involutions as well:

$$T_{L2}(t, x) = (t, -x) \tag{5.78}$$

$$T_{L3}(t, x) = (-t, -x) \tag{5.79}$$

When I discuss the involutions of the inverse square law in the remainder of this paper, I will assume that we are dealing with the second formulation of the law I have provided since this is the accurate formulation of the inverse square law. A law like the first one I considered would be useful for describing the motion of particles in a world in which objects are electrically or gravitationally attracted to objects on their right, say, but repulsed by objects on their left. Since we don't find ourselves in such a world, the latter inverse square law gives us a more physically interesting differential equation to consider.

5.1.13 Other Equations

The following equations in my analysis have not already been christened with names, to my knowledge, so I will name them as I go:

5.1.13.1 Equation M

Equation M is:

$$x_{ttt} = \frac{x_{tt}^2}{t} - \frac{x_{tt}}{x_t} \tag{5.80}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (1998a, 2000a):

$$T_{M1} = (t, x) = (-t, x) \qquad (5.81)$$

$$T_{M2} = (t, x) = (it, -x) \qquad (5.82)$$

$$T_{M3} = (t, x) = (-it, -x) \qquad (5.83)$$

Only the first of these symmetries is an involution.

5.1.13.2 Equation N

Equation N is:

$$x_{ttt} = x_{tt}(1 - x_{tt}) \qquad (5.84)$$

It has the following non-trivial discrete symmetries, according to *Hydon* (2001):

$$T_{N1}(t, x, x_t) = (-t, \frac{1}{2}t^2 - x, x_t - t) \qquad (5.85)$$

$$T_{N2}(t, x, x_t) = \left(-x_t, x - tx_t + \frac{1}{2}x_t^2, t - x_t\right) \qquad (5.86)$$

Only the first of these symmetries, T_{N1}, is an involution.

5.1.13.3 Equation O

Equation O is:

$$x_{ttt} = \frac{2x_{ttt}}{x}(1 - x_t) \qquad (5.87)$$

It has the following non-trivial discrete symmetries, according to *Hydon* (2001):

$$T_{O1}(t, x) = (-t, -x) \tag{5.88}$$

$$T_{O2}(t, x) = \left(-\frac{1}{t}, \frac{x}{t^2}\right) \tag{5.89}$$

Both of these symmetries are involutions.

5.1.13.4 Equation P

Equation P is:

$$x_{tt}^2 = \left(\frac{x_t}{t} - e^x\right)^2 \tag{5.90}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (2001):

$$T_{P1}(t, x) = (-t, x) \tag{5.91}$$

$$T_{P2}(t, x) = \left(\frac{1}{t}, x + 4\ln|t|\right) \tag{5.92}$$

Both of these symmetries are involutions.

5.1.13.5 Equation Q

Equation Q is:

$$x_{tt} = \tan x_t \tag{5.93}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (1998a):

$$T_{Q1,k}(t, x) = (t, x + k\pi t) \tag{5.94}$$

$$T_{Q2,k}(t, x) = (t, -x + k\pi t) \tag{5.95}$$

Where k is any integer. Both sets of symmetries, excluding $T_{Q1,k=0}$, are non-trivial involutions.

5.1.13.6 Equation R

Equation R is:

$$x_{tt} = \frac{x_t}{t} + \frac{4x^2}{t^3} \tag{5.96}$$

It has the following non-trivial discrete symmetries, according to *Hydon* (1998a, 1999):

$$T_{R1}(t, x) = (-t, -x) \tag{5.97}$$

$$T_{R2}(t, x) = \left(\frac{1}{t}, \frac{x}{t^2}\right) \tag{5.98}$$

$$T_{R3}(t, x) = \left(-\frac{1}{t}, -\frac{x}{t^2}\right) \tag{5.99}$$

All three of these are involutions.

The equation can also be written as:

$$\frac{d^2r}{ds^2} = 4r^2 + r \tag{5.100}$$

With $r = \frac{x}{t}$ and $s = \ln|t|$. In this form, the equation has the following non-trivial discrete symmetries:

$$T_{R4}(r, s) = (r, -s) \tag{5.101}$$

$$T_{R5}(r, s) = \left(-r - \frac{1}{4}, is\right) \tag{5.102}$$

$$T_{R6}(r, s) = \left(-r - \frac{1}{4}, -is\right) \tag{5.103}$$

Only the first of these symmetries is an involution. This is not particularly surprising since $T_{R1}(r, s) = (r, s)$ and both T_{R2} and T_{R3} induce T_{R4}'s transformation of r and s.

5.2 Analysis

Tables 1 and 2 summarize the results of the previous section, listing all of the non-trivial involutive point symmetries for each of the differential equations discussed in the previous section. Before using the information on these tables to determine what time reversal would look like if some subset of differential equations examined above served as the fundamental laws of our world, here's a quick review of how to apply my "third way" to determine what the time reversal operator is:

Step 1: Find a set of differential equations that seem the best available candidates for the fundamental physical laws governing the world.

Step 2: Determine all of the discrete symmetries under which these differential equations are invariant.

Step 3: Apply the intuitive criteria for a time reversal operator to rule out those discrete symmetries that cannot be time reversal operators. This means:

Table 5.1: Involutive Candidate T Operators

Name	Equation	Coordinate Transformation
T_{A1}	Black-Scholes	$T_{A1}(x,t,u) = (e^{2Dt-log(x)}, t, u)$
T_{B1}	Burgers'	$T_{B1}(x,t,u) = (-x, t, -u)$
T_C	Spherical Burgers'	$T_C(x,t,u) = (-x, t, -u)$
$T_{D4,\lambda=0}$	TD Burgers'	$T_{D4,\lambda=0}(x,t,u) = (-x, t, -u)$
T_{E1}	Chazy	$T_{E1}(t,x) = (-t, -x)$
T_{E2}	Chazy	$T_{E2}(t,x) = (-\frac{1}{t}, t^2 x + 6t)$
T_{E3}	Chazy	$T_{E3}(t,x) = (\frac{1}{t}, -(t^2 x + 6t))$
T_{F1}	EPD	$T_{F1}(x,t,u) = (-x, t, u)$
T_{F2}	EPD	$T_{F2}(x,t,u) = (x, -t, u)$
T_{F3}	EPD	$T_{F3}(x,t,u) = (x, t, -u)$
T_{F4}	EPD	$T_{F4}(x,t,u) = (-x, -t, u)$
T_{F5}	EPD	$T_{F5}(x,t,u) = (x, -t, -u)$
T_{F6}	EPD	$T_{F6}(x,t,u) = (-x, t, -u)$
T_{F7}	EPD	$T_{F7}(x,t,u) = (-x, -t, -u)$
T_{F8}	EPD	$T_{F8}(x,t,u) = (\frac{x}{t^2-x^2}, \frac{t}{t^2-x^2}, u)$
T_{F9}	EPD	$T_{F9}(x,t,u) = (\frac{-x}{t^2-x^2}, \frac{t}{t^2-x^2}, u)$
T_{F10}	EPD	$T_{F10}(x,t,u) = (\frac{x}{t^2-x^2}, \frac{-t}{t^2-x^2}, u)$
T_{F11}	EPD	$T_{F11}(x,t,u) = (\frac{x}{t^2-x^2}, \frac{t}{t^2-x^2}, -u)$
T_{F12}	EPD	$T_{F12}(x,t,u) = (\frac{-x}{t^2-x^2}, \frac{-t}{t^2-x^2}, u)$
T_{F13}	EPD	$T_{F13}(x,t,u) = (\frac{x}{t^2-x^2}, \frac{-t}{t^2-x^2}, -u)$
T_{F14}	EPD	$T_{F14}(x,t,u) = (\frac{-x}{t^2-x^2}, \frac{t}{t^2-x^2}, -u)$
T_{F15}	EPD	$T_{F15}(x,t,u) = (\frac{-x}{t^2-x^2}, \frac{-t}{t^2-x^2}, -u)$
T_{G2}	Gas	$T_{G2}(s,y,v,p,q,\phi) = (-s, y, -v, p, q, \phi)$
T_{G3}	Gas	$T_{G3}(s,y,v,p,q,\phi) = (s, -y, -v, p, q, -\phi)$
T_{H1}	Harry-Dym	$T_{H1}(x,t,u) = (x, -t, -u)$
T_{H2}	Harry-Dym	$T_{H2}(x,t,u) = (-x, t, -u)$
T_{H3}	Harry-Dym	$T_{H3}(x,t,u) = (-x, -t, u)$
T_{H4}	Harry-Dym	$T_{H4}(x,t,u) = (-\frac{1}{x}, t, \frac{u}{x^2})$
T_{H5}	Harry-Dym	$T_{H5}(x,t,u) = (\frac{1}{x}, t, \frac{-u}{x^2})$
T_{H6}	Harry-Dym	$T_{H6}(x,t,u) = (-\frac{1}{x}, -t, \frac{-u}{x^2})$
T_{H7}	Harry-Dym	$T_{H7}(x,t,u) = (\frac{1}{x}, -t, \frac{u}{x^2})$
$T_{I1,C=-1}$	Heat	$T_{I1,C=-1}(x,t,u) = (-x, t, -u)$
$T_{I1,C=1}$	Heat	$T_{I1,C=1}(x,t,u) = (-x, t, u)$
$T_{I2,C=-1}$	Heat	$T_{I2,C=-1}(x,t,u) = (x, t, -u)$
T_{J1}	Toda	$T_{J1}(x,t,u) = (x, -t, u)$
T_{J2}	Toda	$T_{J2}(x,t,u) = (-x, t, -u)$
T_{J3}	Toda	$T_{J3}(x,t,u) = (-x, -t, -u)$

Table 5.2: Involutive Candidate T Operators (Continued)

Name	Equation	Coordinate Transformation
T_{K1}	Free	$T_{K1}(t,x) = (-t, x)$
T_{K2}	Free	$T_{K2}(t,x) = (t, -x)$
T_{K3}	Free	$T_{K3}(t,x) = (-t, -x)$
T_{K4}	Free	$T_{K4}(t,x) = (x, t)$
T_{K5}	Free	$T_{K5}(t,x) = (-x, -t)$
T_{K6}	Free	$T_{K6}(t,x) = (\frac{t}{x}, \frac{1}{x})$
T_{K7}	Free	$T_{K7}(t,x) = (-\frac{t}{x}, \frac{1}{x})$
T_{K8}	Free	$T_{K8}(t,x) = (\frac{1}{t}, \frac{x}{t})$
T_{K9}	Free	$T_{K9}(t,x) = (\frac{1}{t}, -\frac{x}{t})$
T_{L1}	Inverse Square	$T_{K1}(t,x) = (-t, x)$
T_{L2}	Inverse Square	$T_{K2}(t,x) = (t, -x)$
T_{L3}	Inverse Square	$T_{K3}(t,x) = (-t, -x)$
T_{M1}	M	$T_{M1}(t,x) = (-t, x)$
T_{N1}	N	$T_{N1}(t,x,x_t) = (-t, \frac{1}{2}t^2 - x, x_t - t)$
T_{O1}	O	$T_{O1}(t,x) = (-t, -x)$
T_{O2}	O	$T_{O2}(t,x) = (-\frac{1}{t}, \frac{x}{t^2})$
T_{P1}	P	$T_{P1}(t,x) = (-t, x)$
T_{P2}	P	$T_{P2}(t,x) = (\frac{1}{t}, x + 4\ln|t|)$
$T_{Q1,k \neq 0}$	Q	$T_{Q1,k}(t,x) = (t, x + k\pi t)$
$T_{Q2,k}$	Q	$T_{Q2,k}(t,x) = (t, -x + k\pi t)$
T_{R1}	R	$T_{R1}(t,x) = (-t, -x)$
T_{R2}	R	$T_{R2}(t,x) = (\frac{1}{t}, \frac{x}{t^2})$
T_{R3}	R	$T_{R3}(t,x) = (-\frac{1}{t}, -\frac{x}{t^2})$
T_{R4}	R	$T_{R4}(r,s) = (r, -s)$

A: Rule out all non-involutions.

B: Determine which of the variables represent spatial and temporal coordinates, and rule out all symmetries that do not leave the spatial coordinates unchanged while changing the temporal coordinates.

C: Rule out all symmetries that do not satisfy the consistency condition laid out in the previous chapter.

Step 4: Any remaining discrete symmetries can be properly called "time reversal".

So, having investigated a number of differential equations and determined all of the involutions under which these equations are invariant, there are essentially three more important pieces of information necessary to determine what the time reversal operator should look like: first, we need to determine which of these differential equations provide the best available candidates for the fundamental physical laws governing the world; secondly, we need to determine which variables appealed to by these fundamental equations are interpretively identical (i.e. which variables obtain their values from exactly the same measurement processes); and finally, we need to know which of these variables is time and which is position. Unfortunately, on the first count, none of these differential equations are the sort of things that contemporary physicists would call our best candidates for physical laws governing the world since none appear in the most successful current fundamental physical theories (e.g. quantum field theory, general relativity, statistical mechanics, etc.). The Black-Scholes equation, for instance, is principally of interest to economists as it allows one to determine the price of an option over time. The Chazy equation is of interest because of certain mathematical properties it has but as of yet is of no physical interest. The differential equations listed here as equations L through Q are likewise of interest mathematically but, to my knowledge, carry no physical significance. Some of the differential

equations above, namely the heat equation, the inverse square law, and Burgers' equation do come from physics, but to call them the best available candidates for the fundamental physical laws governing our world would strain credulity.

In the absence of any differential equations in my analysis that seem the best available candidates for the fundamental physical laws governing *our* world, then, we cannot, given the tools above, answer the crucial question of what time reversal *in our own world* must be; however, we can answer the interesting question of what the time reversal operator is in worlds fundamentally governed by the above differential equations (that is, worlds in which the set or some subset of the above differential equations provides the set of best available candidates for the fundamental physical laws governing the world). By turning our focus from the actual world to these other possible worlds, we have an added three degrees of flexibility that allow us to posit the information necessary for applying my method: 1) we can stipulate which of the above differential equations we take to be fundamental at the world in question, 2) we can stipulate which of the independent and dependent variables in each differential equation are interpretively identical to the independent and dependent variables appearing in the other fundamental differential equations, and 3) we can stipulate which of the independent variables corresponds to time and which of the dependent or independent variables corresponds to a spatial coordinate.

I will proceed to analyze a number of these possible worlds and adopt the following naming convention for these worlds: I will label worlds as $W = \langle L; I; T; X \rangle$, where L is the set of differential equations that fundamentally govern the world, I is the set of interpretive identity statements among the variables appearing in L, and T and X are the sets of variables appearing in L which representing temporal and spatial coordinates respectively. I will append to each law in L a numerical subscript so that it is clear which variables are interpretively identical to which other variables, and I will translate $A = B$ as "A is interpretively identical to B".

With these preliminaries out of the way, let's turn to several worlds of interest. I will limit my attention in the remainder of this section to four cases of interest: first, a world in which time reversal works as one might intuitively expect; secondly, a world where time reversal necessitates changes in non-temporal variables; third, a world where time reversal may transform the time coordinate in an unusual way; and finally, a world that may seem closest to the actual world of all of the possible worlds here discussed.

5.2.1 An Analysis of W_1

Take $W_1 = \langle \text{EPD}_1, \text{Toda}_2; x_1 = x_2, t_1 = t_2; t_1, t_2; x_1, x_2 \rangle$. This is the world fundamentally governed by the Euler-Poisson-Darboux equation and the Toda equation where the x coordinates refer to spatial location and the t coordinates refer to time. We assume that the two fields generated by these different differential equations, u_1 and u_2, refer to different physical features of the world so that the two laws are not inconsistent with one another. According to Table 1, since the two fundamental differential equations in W_1 are the EPD and Toda equations, the candidate time reversal operators are the following transformations:

$$T_{F1}(x_1, t_1, u_1) = (-x_1, t_1, u_1) \tag{5.104}$$

$$T_{F2}(x_1, t_1, u_1) = (x_1, -t_1, u_1) \tag{5.105}$$

$$T_{F3}(x_1, t_1, u_1) = (x_1, t_1, -u_1) \tag{5.106}$$

$$T_{F4}(x_1, t_1, u_1) = (-x_1, -t_1, u_1) \tag{5.107}$$

$$T_{F5}(x_1, t_1, u_1) = (x_1, -t_1, -u_1) \tag{5.108}$$

$$T_{F6}(x_1, t_1, u_1) = (-x_1, t_1, -u_1) \tag{5.109}$$

$$T_{F7}(x_1, t_1, u_1) = (-x_1, -t_1, -u_1) \tag{5.110}$$

$$T_{F8}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, u \right) \tag{5.111}$$

$$T_{F9}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, u \right) \tag{5.112}$$

$$T_{F10}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, u \right) \tag{5.113}$$

$$T_{F11}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, -u \right) \tag{5.114}$$

$$T_{F12}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, u \right) \tag{5.115}$$

$$T_{F13}(x, t, u) = \left(\frac{x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, -u \right) \tag{5.116}$$

$$T_{F14}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{t}{t^2 - x^2}, -u \right) \tag{5.117}$$

$$T_{F15}(x, t, u) = \left(\frac{-x}{t^2 - x^2}, \frac{-t}{t^2 - x^2}, -u \right) \tag{5.118}$$

$$T_{J1}(x_2, t_2, u_2) = (x_2, -t_2, u_2) \tag{5.119}$$

$$T_{J2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.120}$$

$$T_{J3}(x_2, t_2, u_2) = (-x_2, -t_2, -u_2) \tag{5.121}$$

We can now begin to winnow down the field of time reversal operator candidates by weeding out all of the transformations that result in changes to the spatial coordinates, as suggested in part B of step 3 described above. This leaves us with:

$$T_{F2}(x_1, t_1, u_1) = (x_1, -t_1, u_1) \tag{5.122}$$

$$T_{F3}(x_1, t_1, u_1) = (x_1, t_1, -u_1) \tag{5.123}$$

$$T_{F5}(x_1, t_1, u_1) = (x_1, -t_1, -u_1) \tag{5.124}$$

$$T_{J1}(x_2, t_2, u_2) = (x_2, -t_2, u_2) \tag{5.125}$$

as the candidates for time reversal operator. The only time reversal candidate under which the Toda equation is invariant, we see, is T_{J1}. We can now apply part C of

step 3 above by using the consistency constraint to rule out T_{F3} as a suitable time reversal operator since it transforms the time coordinate t_1 in a way that T_{J1} does not. This leaves us with three candidates for the time reversal operator:

$$T_{F2}(x_1, t_1, u_1) = (x_1, -t_1, u_1) \tag{5.126}$$

$$T_{F5}(x_1, t_1, u_1) = (x_1, -t_1, -u_1) \tag{5.127}$$

$$T_{J1}(x_2, t_2, u_2) = (x_2, -t_2, u_2) \tag{5.128}$$

This tells us that time reversal operator in W_1 works as follows on the following coordinates:

$$T(x) = x \tag{5.129}$$

$$T(t) = -t \tag{5.130}$$

$$T(u_2) = u_2 \tag{5.131}$$

But what about $T(u_1)$? Sine both T_{F2} and T_{F5} are time reversal operator candidates and since these two transformations only differ with respect to their action on u_1, the answer is underdetermined, and there is no fact of the matter (at the moment, at least) as to how time reversal acts on u_1 (though it is worth noting that it must either leave u_1 unchanged or flip its sign). This underdetermination may seem problematic since my analysis does not provide a definitive answer to the question of how time reversal transforms this variable, but remember that I have stipulated nothing about the nature of the field u_1 yet, so there may be some further fact about the nature of u_1 that can prove decisive, such as some as-yet-undiscovered law governing W_1 whose discrete symmetries, by consistency constraints, may help us determine how

T transforms u_1. In a way, then, this underdetermined result is of great value to the physicist since it may suggest new avenues of research into basic physical laws that may not have occurred to the physicist otherwise.

But setting aside all problems with u_1, the analysis of time reversal in W_1 is fairly straightforward: distances of Δx will remain unchanged by time reversal while a duration of Δt will change to $-\Delta t$, giving us a time reversal operator like the ones proposed by Horwich and Albert. That is, if the inhabitants of W_1 were to watch a movie and its time-reversed counterpart, they would see one playing "backwards" at the same speed the other was playing "forwards". What's more, the field u_2, while a function of t, is left unaffected by the time reversal process, as many fundamental fields, at least according to Albert, are wont to do. So my account provides us, at least in this case, with a relatively uninteresting time reversal operator for W_1.

5.2.2 An Analysis of W_2

Take $W_2 = \langle$SphericalBurgers'$_1$,Harry-Dym$_2$,Heat$_3; x_1 = x_2 = x_3, t_1 = t_2 = t_3; x_1, x_2, x_3; t_1, t_2, t_3\rangle$. This is the world fundamentally governed by the Spherical Burgers' equation, the Harry-Dym equation, and the heat equation where the x coordinates, which are all interpretively identical with one another, refer to time while the t coordinates, which are all interpretively identical with one another, refer to spatial position. It's worth noting here how strange this assignment would be for those who agree with *Callender* (2011) and *Skow* (2007) and think that the time variable must be the variable for which the equation's initial value problem is well-behaved, but, for the moment, assume their view is wrong and that we have very good evidence to believe that the x coordinates really represent time in these equations.[4] Again, we take the three fields u_1, u_2, and u_3 to represent three different physical quantities but make no further assertions about these fields. According to Table 1, the following are

---

[4]Similar strangeness will occur when we turn next to W_3 and consider a world in which time is a dependent rather than an independent variable in one of our laws.

the candidates for the time reversal operator:

$$T_C(x_1, t_1, u_1) = (-x_1, t_1, -u_1) \tag{5.132}$$

$$T_{H1}(x_2, t_2, u_2) = (x_2, -t_2, -u_2) \tag{5.133}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.134}$$

$$T_{H3}(x_2, t_2, u_2) = (-x_2, -t_2, u_2) \tag{5.135}$$

$$T_{H4}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, t_2, \frac{u_2}{x_2^2}\right) \tag{5.136}$$

$$T_{H5}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, t_2, \frac{-u_2}{x_2^2}\right) \tag{5.137}$$

$$T_{H6}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, -t_2, \frac{-u_2}{x_2^2}\right) \tag{5.138}$$

$$T_{H7}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, -t_2, \frac{u_2}{x_2^2}\right) \tag{5.139}$$

$$T_{I1,C=-1}(x_3, t_3, u_3) = (-x_3, t_3, -u_3) \tag{5.140}$$

$$T_{I1,C=1}(x_3, t_3, u_3) = (-x_3, t_3, u_3) \tag{5.141}$$

$$T_{I2,C=-1}(x_3, t_3, u_3) = (x_3, t_3, -u_3) \tag{5.142}$$

We can now begin to winnow down the field of time reversal operator candidates by weeding out all of the transformations that result in changes to the spatial coordinates (which, remember, in W_2 are the t-coordinates) or which don't result in changes to the temporal coordinates (which, remember, in W_2 are the x-coordinates), as suggested in part B of step 3 described above. This leaves us with:

$$T_C(x_1, t_1, u_1) = (-x_1, t_1, -u_1) \tag{5.143}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.144}$$

$$T_{H4}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, t_2, \frac{u_2}{x_2^2}\right) \tag{5.145}$$

$$T_{H5}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, t_2, \frac{-u_2}{x_2^2}\right) \tag{5.146}$$

$$T_{I1,C=-1}(x_3, t_3, u_3) = (-x_3, t_3, -u_3) \tag{5.147}$$

$$T_{I1,C=1}(x_3, t_3, u_3) = (-x_3, t_3, u_3) \tag{5.148}$$

as the candidates for time reversal operator. The only time reversal candidate under which the Spherical Burgers' equation is invariant is T_C, and it remains after this first winnowing-down. We can now apply part C of step 3 above, using the consistency constraint to rule out T_{H4} and T_{H5} as a suitable time reversal operators since they transform the time coordinate x_2 in a way that T_C does not transform x_1. This leaves us with four candidates for the time reversal operator:

$$T_C(x_1, t_1, u_1) = (-x_1, t_1, -u_1) \tag{5.149}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.150}$$

$$T_{I1,C=-1}(x_3, t_3, u_3) = (-x_3, t_3, -u_3) \tag{5.151}$$

$$T_{I1,C=1}(x_3, t_3, u_3) = (-x_3, t_3, u_3) \tag{5.152}$$

This tells us that the time reversal operator in W_2 works as follows on the following coordinates (here, using x and t to refer to position and time across equations respectively):

$$T(x) = x \tag{5.153}$$

$$T(t) = -t \tag{5.154}$$

$$T(u_1) = -u_1 \tag{5.155}$$

$$T(u_2) = -u_2 \tag{5.156}$$

Again, we have a field whose behavior under time reversal is underdetermined (u_3 in this case), but more interesting is the fact that time reversal flips the signs of both u_1 and u_2; that is, the time reversal operator has a non-trivial effect on these coordinates. The switching of u_1's sign under time reversal may not be particularly surprising since T_C was the only time reversal candidate for the spherical Burgers' equation, but surprisingly the Harry-Dym equation, which had seven time reversal candidates before winnowing the number down by applying my methodology, has only one remaining discrete symmetry which can be a candidate for time reversal. The consistency condition here shows its true power since it is this condition that ruled out three alternatives to T_{H2}.

What is interesting about W_2 is that we know how u_1 and u_2 both must transform under T in order for the fundamental laws of W_2 to be time reversal invariant even though we know nothing about the physical quantities that u_1 and u_2 represent. Intuitive time reversal theorists like Albert may take this to provide us with the information that values of these fields (unlike u_1 in W_1) cannot be taken as fundamental properties of instantaneous states but must rather be taken as relations among a number of instantaneous states, though I make no such claim. What does seem relevant, however, is that we've found a world where fundamental time reversal invariance might just as well be called "fundamental time *and u_1 and u_2* reversal invariance" in the same way that *Greaves* (2010) suggests that time reversal invariance

may just as well be called "time *and charge* reversal invariance". My account thus allows for the kind of meaningful empirical discoveries about time reversal invariance that seem crucial to the projects of philosophers like Greaves, and so I take the results of my method in cases like W_2 to give those interested in conducting such projects reason to take on my view of time reversal.

But again, in some respects the results in W_2 are quite expected. The inhabitants of W_1, much like the inhabitants of W_2, will see a movie and its time-reversed counterpart as playing at the same speed but in opposite directions. Both take the time reversal operator to flip the sign of the temporal coordinate while leaving the spatial coordinate unchanged. In the next section, however, time reversal will seem a bit weirder.

5.2.3 An Analysis of W_3

Take $W_3 = \langle \text{Chazy}_1, \text{Harry-Dym}_2, \text{P}_3, \text{R}_4; t_1 = x_2 = t_3 = t_4; t_1, x_2, t_3, t_4; t_2 \rangle$. This is the world fundamentally governed by the Chazy equation, the Harry-Dym equation, and Equations P and R where the t_1, x_2, t_3, and t_4, which are all interpretively identical to one another, refer to time while t_2 alone refers to spatial position. Again, we take the fields x_1, u_2, x_3, and x_4 to represent three different physical quantities but make no further assertions about these fields. According to Tables 1 and 2, the following are the candidates for the time reversal operator:

$$T_{E1}(t_1, x_1) = (-t_1, -x_1) \tag{5.157}$$

$$T_{E2}(t_1, x_1) = \left(-\frac{1}{t_1}, t_1^2 x_1 + 6t_1\right) \tag{5.158}$$

$$T_{E3}(t_1, x_1) = \left(\frac{1}{t_1}, -(t_1^2 x_1 + 6t_1)\right) \tag{5.159}$$

$$T_{H1}(x_2, t_2, u_2) = (x_2, -t_2, -u_2) \tag{5.160}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.161}$$

$$T_{H3}(x_2, t_2, u_2) = (-x_2, -t_2, u_2) \tag{5.162}$$

$$T_{H4}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, t_2, \frac{u_2}{x_2^2}\right) \tag{5.163}$$

$$T_{H5}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, t_2, \frac{-u_2}{x_2^2}\right) \tag{5.164}$$

$$T_{H6}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, -t_2, \frac{-u_2}{x_2^2}\right) \tag{5.165}$$

$$T_{H7}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, -t_2, \frac{u_2}{x_2^2}\right) \tag{5.166}$$

$$T_{P1}(t_3, x_3) = (-t_3, x_3) \tag{5.167}$$

$$T_{P2}(t_3, x_3) = \left(\frac{1}{t_3}, x_3 + 4\ln|t_3|\right) \tag{5.168}$$

$$T_{R1}(t_4, x_4) = (-t_4, -x_4) \tag{5.169}$$

$$T_{R2}(t_4, x_4) = \left(\frac{1}{t_4}, \frac{x_4}{t_4^2}\right) \tag{5.170}$$

$$T_{R3}(t_4, x_4) = \left(-\frac{1}{t_4}, -\frac{x_4}{t_4^2}\right) \tag{5.171}$$

We can now begin to winnow down the field of time reversal operator candidates by weeding out all of the transformations that result in changes to the spatial coordinates (which, remember, in W_3, is just t_2) or which don't result in changes to the temporal coordinates (which, remember, in W_3 are t_1, x_2, t_3, and t_4), as suggested in part B of step 3 described above. This leaves us with:

$$T_{E1}(t_1, x_1) = (-t_1, -x_1) \tag{5.172}$$

$$T_{E2}(t_1, x_1) = \left(-\frac{1}{t_1}, t_1^2 x_1 + 6t_1\right) \tag{5.173}$$

$$T_{E3}(t_1, x_1) = \left(\frac{1}{t_1}, -(t_1^2 x_1 + 6t_1)\right) \tag{5.174}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.175}$$

$$T_{H4}(x_2, t_2, u_2) = \left(-\frac{1}{x_2}, t_2, \frac{u_2}{x_2^2}\right) \tag{5.176}$$

$$T_{H5}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, t_2, \frac{-u_2}{x_2^2}\right) \tag{5.177}$$

$$T_{P1}(t_3, x_3) = (-t_3, x_3) \tag{5.178}$$

$$T_{P2}(t_3, x_3) = \left(\frac{1}{t_3}, x_3 + 4\ln|t_3|\right) \tag{5.179}$$

$$T_{R1}(t_4, x_4) = (-t_4, -x_4) \tag{5.180}$$

$$T_{R2}(t_4, x_4) = \left(\frac{1}{t_4}, \frac{x_4}{t_4^2}\right) \tag{5.181}$$

$$T_{R3}(t_4, x_4) = \left(-\frac{1}{t_4}, -\frac{x_4}{t_4^2}\right) \tag{5.182}$$

as the candidates for time reversal operator. There are many more transformations left after this step of my analysis than in the previous two cases, and we are left with two candidate time reversal operators for equation P: T_{P1} and T_{P2}. Let's assume that T_{P1} is the "true" time reversal operator. If this is so, then the consistency constraints leave us with the following set of transformations as time reversal operator candidates:

$$T_{E1}(t_1, x_1) = (-t_1, -x_1) \tag{5.183}$$

$$T_{H2}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.184}$$

$$T_{P1}(t_3, x_3) = (-t_3, x_3) \tag{5.185}$$

$$T_{R1}(t_4, x_4) = (-t_4, -x_4) \tag{5.186}$$

If we take T_{P2} as the correct time reversal operator, then consistency constraints leave us with the follow set of transformations as time reversal operator candidates:

$$T_{E3}(t_1, x_1) = \left(\frac{1}{t_1}, -(t_1^2 x_1 + 6t_1) \right) \tag{5.187}$$

$$T_{H5}(x_2, t_2, u_2) = \left(\frac{1}{x_2}, t_2, \frac{-u_2}{x_2^2} \right) \tag{5.188}$$

$$T_{P2}(t_3, x_3) = \left(\frac{1}{t_3}, x_3 + 4\ln|t_3| \right) \tag{5.189}$$

$$T_{R2}(t_4, x_4) = \left(\frac{1}{t_4}, \frac{x_4}{t_4^2} \right) \tag{5.190}$$

There are, then, essentially two candidates for the time reversal operator. The first, T_1, works as follows (again, taking x, without subscripts, to now refer to the position coordinate and t to now refer to the temporal coordinate):

$$T_1(x) = x \tag{5.191}$$

$$T_1(t) = -t \tag{5.192}$$

$$T_1(x_1) = -x_1 \tag{5.193}$$

$$T_1(u_2) = -u_2 \tag{5.194}$$

$$T_1(x_3) = x_3 \tag{5.195}$$

$$T_1(x_4) = -x_4 \tag{5.196}$$

The second, T_2, works as follows:

$$T_2(x) = x \tag{5.197}$$

$$T_2(t) = \frac{1}{t} \tag{5.198}$$

$$T_2(x_1) = -(t^2 x_1 + 6t) \tag{5.199}$$

$$T_2(u_2) = \frac{-u_2}{t^2} \tag{5.200}$$

$$T_2(x_3) = x_3 + 4\ln|t| \tag{5.201}$$

$$T_2(x_4) = \frac{x_4}{t^2} \tag{5.202}$$

Several things are worth remarking on at this point. First, whether one takes T_1 or T_2 to be the "true" or "best" time reversal operator in W_3, we are left with a time reversal operator that determinately transforms all of the coordinates we have examined here, including x_1, u_2, x_3, and x_4. We are also obviously left with two distinct ways the world could be time reversal invariant. But perhaps the most interesting result about W_3 is provided by T_2, a strange-looking time reversal operator which, odd though it is, still makes a kind of sense as a time reversal operator. As with T_1, displacements of Δx remain unchanged by T_2, but T_2 transforms the temporal coordinates in such a way that the distance between two events (say, t_0 and t_1 with displacement $\Delta t = t_0 - t_1$) changes to $-\frac{\Delta t}{t_0 t_1}$. So, were the inhabitants of W_3 to watch a movie, its T_1 counterpart, and its T_2 counterpart, they would note that, while the T_1 and T_2 counterparts were running backwards relative to the initial movie, some segments of the T_2 counterpart would move more slowly than those corresponding segments in the T_1 counterparts, and others would move more quickly.

T_2, while an entirely different transformation than what we've seen previously, thus has some important similarities with the more intuitively plausible T_1 such as the way it inverts a sequence of events, but by adding in an additional time-dependent scaling factor, the overall transformation looks quite different than the one provided

by T_1. I believe this shows an additional virtue of my account of time reversal in that it suggests interesting new ways the world could be time reversal invariant. Of course, W_3 on its own is not necessarily such a world, but when conducting further empirical research, scientists in W_3 should be on the lookout for differential equations whose symmetries would, via consistency, establish either T_1 or T_2 as the time reversal operator in W_3. Intuitive time reversal theorists would likely balk at a time reversal operator like T_2, but I see no reason to do so without any further information; again, the fundamental laws are invariant under this transformation, and it inverts sequences of events as required, so there seems no principled reason to reject fascinating and bizarre transformations like T_2 as candidate time reversal operators out of hand.

5.2.4 An Analysis of W_4

Take $W_4 = \langle\text{Burgers'}_1,\text{Heat}_2,\text{Inverse-Square}_3,\text{Free}_4; x_1 = x_2 = x_3 = x_4, t_1 = t_2 = t_3 = t_4; t_1, t_2, t_3, t_4; x_1, x_2, x_3, x_4\rangle$. W_4 is closer to the actual world than many of the possible worlds we have examined up to this point. Its fundamental equations are the one-dimensional Burgers' equation, which is used to model waves in acoustics and hydrodynamics; the one-dimensional heat equation, which gives the temperature of a material as a function of both time and position; the the one-dimensional inverse square law equation, which allows one to determine the distance between a Newtonian free particle and a source of gravitational attraction as a function of time; and the equation for the free particle, which allows one to determine the motion of a Newtonian particle in the absence of any external forces. Clearly, these differential equations are highly idealized and not the sort of equations most physicists would consider to be the most fundamental differential equations that govern the behavior of our universe, and W_4 is doubtless not significantly closer to the actual world than the other worlds considered previously; however, they are the most physically significant of the differential equations we have to work with, and so an analysis of these

differential equations is still of some physical interest. The candidate time reversal operators for W_4 are:

$$T_{B1}(x_1, t_1, u_1) = (-x_1, t_1, -u_1) \tag{5.203}$$

$$T_{I1,C=-1}(x_2, t_2, u_2) = (-x_2, t_2, -u_2) \tag{5.204}$$

$$T_{I1,C=1}(x_2, t_2, u_2) = (-x_2, t_2, u_2) \tag{5.205}$$

$$T_{I2,C=-1}(x_2, t_2, u_2) = (x_2, t_2, -u_2) \tag{5.206}$$

$$T_{L1}(t_3, x_3) = (-t_3, x_3) \tag{5.207}$$

$$T_{L2}(t_3, x_3) = (t_3, -x_3) \tag{5.208}$$

$$T_{L3}(t_3, x_3) = (-t_3, -x_3) \tag{5.209}$$

$$T_{K1}(t_4, x_4) = (-t_4, x_4) \tag{5.210}$$

$$T_{K2}(t_4, x_4) = (t_4, -x_4) \tag{5.211}$$

$$T_{K3}(t_4, x_4) = (-t_4, -x_4) \tag{5.212}$$

$$T_{K4}(t_4, x_4) = (x_4, t_4) \tag{5.213}$$

$$T_{K5}(t_4, x_4) = (-x_4, -t_4) \tag{5.214}$$

$$T_{K6}(t_4, x_4) = \left(\frac{t_4}{x_4}, \frac{1}{x_4}\right) \tag{5.215}$$

$$T_{K7}(t_4, x_4) = \left(-\frac{t_4}{x_4}, \frac{1}{x_4}\right) \tag{5.216}$$

$$T_{K8}(t_4, x_4) = \left(\frac{1}{t_4}, \frac{x_4}{t_4}\right) \tag{5.217}$$

$$T_{K9}(t_4, x_4) = \left(\frac{1}{t_4}, -\frac{x_4}{t_4}\right) \tag{5.218}$$

We can now begin to winnow down the field of time reversal operator candidates by weeding out all of the transformations that result in changes to the spatial coordinates or which don't result in changes to the temporal coordinates, as suggested in part B of step 3 described above. This leaves us with only two candidate time reversal

operators:

$$T_{L1}(t_3, x_3) = (-t_3, x_3) \tag{5.219}$$

$$T_{K1}(t_4, x_4) = (-t_4, x_4) \tag{5.220}$$

We have eliminated, then, all of the discrete symmetries under which both the heat equation and Burgers' equation are invariant from contention. The only conclusion, then, is that the fundamental laws of physics in W_4 are not all invariant under time reversal. The best remaining candidate for a time reversal operator may be T_1, which acts as follows:

$$T_1(x) = x \tag{5.221}$$

$$T_1(t) = -t \tag{5.222}$$

but even with this in place, the conclusion about the time reversal non-invariance of the physical laws still follows. There may be space-time reversals under which the laws are invariant, but there is no time reversal operator one could formulate that would allow this set of laws to be time reversal invariant.

What are we to make of this result? Again, in one sense, it's not particularly surprising to find that two differential equations, one of which comes from fluid dynamics and the other of which is dissipative[5], are not time reversal invariant while there seems a decent time reversal operator in the Newtonian differential equations that would make said equation time reversal invariant.[6] But the more important

---

[5]There is no way of inferring a unique set of initial conditions for the heat equation using only the temperature distribution in a material at a particular time. See page 44 of *Earman* (1986) for more details.

[6]It's especially unsurprising when one considers that T_1 is exactly the sort of time reversal operator that intuitive accounts of time reversal would use when assessing the time reversal invariance

point is that the time reversal non-invariance of W_4 gives us a reason (though it is an extremely weak and defeasible reason) to think that the dynamical laws in the actual world may not all be time reversal invariant. This may not be an earth-shattering revelation to those who think about time reversal invariance since the irreversibility of laws in thermodynamics has given us reason to think that our own physical laws may not be time reversal invariant, and even common experiences such as stirring cream into coffee and being unable to stir it back out again support this intuition; however, it is worth showing that my method can come to similar conclusions when given the appropriate physical laws as resources to work from.

5.3   Conclusion

In concluding this chapter, I should take a moment to compare my analysis to the stances towards time reversal in these four possible worlds adopted by token fictional representatives from both the intuitive and theory-relative camps. I have pointed out along the way already numerous places where my account diverges from intuitive accounts like Albert's, but it's worth remarking on the differences once again. Albert's account would have us take as candidates for time reversal only transformations T such that $T(x) = x$ and $T(t) = -t$. As I've shown above, transformations that satisfy this constraint are among the most common, appearing in W_1 and W_2 (though Albert and others like him would doubtless want more information about how to interpret the fundamental fields in these worlds before agreeing with my analysis of how time reversal transforms these fields). However, my analysis of W_3 suggests that there may be another sort of time reversal (the one I call T_2) which acts on t in a different way. My account challenges intuitive accounts to justify their intuitions that rule out T_2 as a viable time reversal candidate, and, regardless of whether or not intuitive accounts can provide a compelling defense of these intuitions, I believe forcing the intuitive

(or lack thereof) of Newtonian systems.

theorist to engage this challenge will result in progress in better understanding time reversal.

Theory-relative accounts are likewise challenged by my account, though the challenge they face is quite different from the one faced by intuitive accounts. Looking back over the four examples above, it is clear how the "intuitive" criteria I lay down in step 3 parts B and C (as described in the previous section) are essential to paring down the number of candidate time reversal operators for each world, and it is equally clear how essential this paring down of symmetry candidates is for making progress both in determining what the time reversal operator looks like and, based on my analysis in chapter 2, how one might best draw metaphysical conclusions from such a result. The challenge to theory-relative accounts, then, is to explain where the "intuitive" criteria laid down by my account go wrong, for it is these criteria that allow my account to provide more specific details than theory-relative accounts can about the behavior of the time reversal operator in each world. Again, I don't think my challenge reveals a fatal flaw in the reasoning of theory-relative accounts of time reversal, but it is a challenge that must be met if theory-relative accounts wish to show themselves superior to my account.

Though I have applied my method to determine the time reversal operator for various possible worlds governed by small subsets of the differential equations I have catalogued in this chapter, there is still much more work to be done in fleshing out my account of time reversal. Though my analysis of W_1, W_2, and W_3 provides interesting examples of ways the world could be and my analysis of W_4 provides an extremely defeasible reason to think the fundamental laws of physics are not, as physicists like *Carroll* (2010) claim, time reversal invariant, the real test for my account will come when my methodology is applied to a set of differential equations that more aptly represents what we take to be our best available fundamental physical theories. Unfortunately, such an analysis is beyond the scope of this work, but it does provide an

important next step to take from the analysis in this work. In the next chapter, I will turn away from my particular account of time reversal and ask a more general question about the significance of time reversal: even if the fundamental laws of physics are not invariant under time reversal (as my analysis of W_4 may weakly suggest), what kind of interesting facts about the world and our experience of it follow from this fact?

Do Time-Asymmetric Laws Call For Time-Asymmetric Spacetime Structure?

Imagine that you are sitting in front of a television screen that is paused on a particular frame of a movie. You hit a button on your remote control but are not sure whether you have hit "fast forward" or "rewind" because the room around you is dark. As you watch the screen, you see a spoon stir cream into a cup of coffee so that the cream is evening distributed in the coffee by the end of the scene. You infer from this series of events alone that you have hit the "fast forward" button. Alternatively, if you were to see the spoon stir a cup of coffee with the cream evenly distributed in it until all the cream was in one small region of cup, you would infer that you had hit the "rewind" button. In both instances, your inferences are well-justified by physics: thermodynamics and statistical mechanics tell us that we frequently see cream being stirred into coffee but never see it stirred out of coffee, no matter how long or how vigorously we stir.[1] The situation would be quite different if we were watching a video of, say, two billiard balls crashing into one another and then bouncing apart. After hitting "fast-forward" on such a video, we would see the balls crash into each other and then bounce apart, and, had we started at the end and hit "rewind", we would have seen the balls crash into each other and then bounce apart. In short, we can't

---

[1] We must assume, of course, some facts about initial state of the universe in order for these theories to give us this sort of pronouncement.

make any useful inferences about which button we hit if we are watching a video of billiard ball collisions.

If we ask a physicist for more detail about what's going on when the cream is stirred into the cup of coffee, she will tell us a story involving a set of initial conditions and the laws of physics which, jointly, allow us to determine the future state (or the possible future state) of our system (the cup of coffee). If we find, as with our video of the cup of coffee, that it's easy to determine which of two states is the initial state of a physical system and which state is the final state, we may be able to appeal to various temporally asymmetric features of the physical world to explain this phenomenon. For instance, in the case of the cup of coffee, the laws of thermodynamics suggest that the cup of coffee with all of the cream located in some small region is a very special, unlikely state (i.e. a state of low entropy) and the laws of nature will tend to time-evolve this state towards a less special state (i.e. a state of high entropy), namely the state where the cream is evenly distributed throughout the cup. In the case of the billiard balls, however, the initial and final states (with the balls far apart from one another) are equally "special", and so the time reversal invariant (TRI) laws of Newtonian Mechanics tell us that both the "fast forward" and "reversed" sequences of events are physically possible. The laws of statistical mechanics, like Newton's laws, are TRI, and so the temporal irreversibility of thermodynamic processes can be attributed to an asymmetry in the initial and final conditions instead of to an asymmetry in the laws themselves.

However, in some cases, the temporal asymmetry is attributed not to the states related by the physical laws but rather to the laws of physics themselves; that is, there are cases that give us reason to think that some of the fundamental laws of physics are not TRI. It is this second case that is the focus of this paper for the following reason: some authors, most notably *Horwich* (1987) and *Maudlin* (2007), appeal to temporally irreversible laws to justify their claim that there is a temporal

anisotropy in the geometric structure of spacetime itself. This is not a claim about whether spacetime is as the substantivalist claims it is or the relationist claims it is (though Maudlin, at least, is certainly a substantivalist) but rather a claim about what the features of our spacetime, whether substantivalist or relationist, are. Both Horwich and Maudlin take the fact that the laws of physics are not TRI to license our inference to some important fact about the geometric structure of spacetime. But Horwich's and Maudlin's inference from temporally irreversible laws to temporally anisotropic spacetime structure is quick, and it is not clear what they take to justify this inference. Take, for instance, the following claim by *Maudlin* (2007):

> To begin with, the laws of physics as we have them...are not Time Reversal
> Invariant. The discovery that physical processes are not, in *any* sense,
> indifferent to the direction of time is important and well known: it is the
> discovery of the violation of so-called CP invariance, as observed in the
> decay of the neutral K meson. These decays are *not* invariant if one
> changes a right-handed for a left-handed spatial orientation (parity) and
> changes positive to negative charge (charge conjugation). According to
> the CPT theorem, any plausible quantum theory will be invariant under
> parity-plus-charge-conjugation-plus-time-reversal, so the violation of CP
> implies a violation of T. In short, the fundamental laws of physics, as we
> have them, *do* require a temporal orientation on the spacetime manifold.
>
> (117, original emphasis)

In this passage, Maudlin moves from the fact that fundamental laws are not TRI, which is supported by CP violation, to the fact that there must be some sort of temporal anisotropy without making explicit what licenses this inference. The asymmetry in the dynamical laws, then, is taken as evidence for (but is not equivalent to) the kind of spacetime orientation that Maudlin argues for in the rest of this chapter.

A similar passage can be found in *Horwich* (1987):

Once a genuine instance of nomological irreversibility has been identified, it is not hard to justify the inference that time is anisotropic. Suppose there is a physically possible process $ABCD$ whose temporal inverse is impossible. Let $(ABCD)$ designate the process whose temporal orientation is unspecified–merely that B occurs between A and C, and C between B and D. Then a physically necessary condition for the occurrence of $(ABCD)$ is that A is earlier than B. Thus the relation *earlier than* enters into explanations that are fundamental, for we have no deeper account of that necessary condition. In particular, we cannot suppose that the possibility of $(ABCD)$ will be found to depend on its orientation relative to certain other events for in that case the reverse of $ABCD$ would not be physically impossible. (53)

Like Maudlin, Horwich takes the TRI-failure of the fundamental physical laws to give us good reason to believe that spacetime itself is temporally anisotropic. The thesis that underwrites both Horwich's and Maudlin's claims is that temporally irreversible laws license an inference to the existence of temporally asymmetric spacetime structure. But why should this be so? Maudlin does not provide us with an explicit argument, and *Price* (2011) provides us with a way out of Horwich's conclusion:

However, I think this argument overlooks the fact that there will always be a "Machian" reading of the kind of lawlike irreversibility that Horwich has in mind here–simply a law to the effect that all instances of the kind $(ABCD)$ have the same temporal orientation. The Machian law will do the same job of explaining the orientation of any particular $(ABCD)$: the opposite orientation would not match all the other instances. True, it won't

explain why it isn't the case that all instances have the opposite orientation. But even if we grant for the moment that this is a distinct possibility, rather than a notational variant....Horwich's version of the explanation shares an exactly analogous deficit: in his case, there is nothing to explain why we don't find the reverse law, with respect to the temporal anisotropy in question (Why shouldn't it be *later* rather than *earlier* that does the explanatory job, as it were?). (293, original emphasis)

Horwich's argument seems to fail, then, because the Machian alternative to a temporal anisotropy explains the nomic temporal asymmetries as well as Horwich's temporal anisotropy does. If there is a reason to believe that temporally irreversible laws give us a good reason to believe that spacetime itself is temporally anisotropic, then a better argument for this conclusion than the one provided by Horwich is needed. However, we may still contend that Price's explanation here isn't really better than the one Horwich provided; after all, the Machian law is strangely non-local in a way the alternative is not, and for those who favor substantivalist theories of spacetime, we may worry that this proposal pushes us towards relationism. Price's position requires a lengthier defense than what has been provided in this passage. So it seems that further arguments on both sides are required for the debate over what we should conclude about temporal features spacetime given temporally irreversible laws to move forward.

My central claim is that we have no good reason to infer the existence of temporal anistropy in spacetime from only the temporally irreversible laws we have observed. More specifically, I argue that the best arguments in favor of such an inference, one of which draws from work by John Earman and one of which derives from a position advocated by Harvey Brown, both fail. In the second section of the paper, I lay out a promising argument for temporal spacetime anisotropy from temporally irreversible physical laws. In the third section, I discuss, very briefly, the experimental results

130

that give us reason to believe that there are some laws of physics that are temporally irreversible (even if one disagrees with my analysis of the time reversal invariance of laws in the previous chapters of this dissertation), and in the fourth section, I appeal to an argument from Earman to argue that such nomic temporal irreversibility should lead us to introduce some temporally asymmetric structure to our best physical theories.

In the fifth section, however, I argue that such a temporally asymmetric structure need not and should not be understood as a feature of the spacetime structure of our best physical theories, at least at the moment, and is better understood as either a law of nature or a feature of matter fields. This is true regardless of whether we ascribe to substantivalist or relationist theories of spacetime. This conclusion gives us reason to think that Maudlin's and Horwich's inference is misguided. In the sixth section, I consider an alternative argument that relies on the work of *Brown* (2005). Brown's view of spacetime both poses a challenge to my analysis in the fifth section and itself grounds an inference from temporally irreversible laws to temporally anistropic spacetime structure. However, I argue that Brown's view of spacetime is explanatorily inferior to an alternative view proposed by *Janssen* (2009) and, as such, should be rejected. Finally, I draw my conclusions in the final section of the chapter.

6.1   An Argument for Temporal Asymmetry in Spacetime

In the previous section I provided Price's response to Horwich's argument in favor of drawing a connection between temporally irreversible laws and a temporally anisotropic spacetime; however, I had no such adequate response to Maudlin's argument, except to say that his argument was far too quick. It seems that he needs a premise along the lines of the following for his conclusion to follow from his premises: "If the fundamental physical laws are not invariant under the time reversal symmetry transformation, then the spacetime required by these fundamental physical laws is

not invariant under the time reversal symmetry transformation." As it stands this premise seems dubious: why should we think that all dynamical asymmetries (that is, asymmetries in the laws of nature) must be asymmetries in the structure of space-time? To be more explicit about the work done by this premise, let's break it into two separate premises and explicitly recast Maudlin's argument as follows:

The Argument for Temporal Asymmetry in Spacetime

1. Not all of our best candidates for fundamental laws of physics are invariant under time reversal, a dynamical symmetry.

2. Any dynamical asymmetry of a physical theory gives us sufficient reason to posit the existence of some similarly asymmetric physical structure.

3. Asymmetric structures posited because of a dynamical asymmetry in the fundamental laws are best interpreted as features of spacetime.

4. The spacetime required by our best candidates for fundamental physical laws is temporally anisotropic. (from 1-3)

Breaking the inference from a dynamical asymmetry to an asymmetry in spacetime into two separate premises in this argument makes it clear exactly what steps must be taken when moving from facts about the laws to facts about spacetime. In the first place, when we discover some sort of dynamical asymmetry in the laws, we should posit the existence of some structure to explain the breaking of this symmetry. We may explain an asymmetry in some law, for instance, by appealing to features of other laws from which the first law may be derived under certain approximations, or we may find that, as pointed out in the introduction, there may be a kind of asymmetry in the states related by the law that gives the appearance of an asymmetry in the laws when, in fact, it is the features of the world that the law depends on that are

asymmetric. Note that this asymmetric structure need not be a *new* structure. For instance, our laws may not be invariant under the 100-dimensional group of rotations mathematically represented by SO(100), but we need not posit some new structure to explain this fact if our theory already appeals to physical structures like tensors and pseudotensors that fail to be invariant under such transformations. However, if we find that the laws fail to be invariant under some symmetry that we currently take all of our physical structures to be invariant under, the second premise of the argument above licenses us to posit some new structure that breaks this symmetry or replace an existing element of the structure with a new element that breaks the symmetry.[2] I examine this premise in detail in the fourth section of this chapter.

But as these examples show, just because we have posited some asymmetric structure to explain or underwrite a particular asymmetry in the dynamical laws does not mean that we always know how to interpret the object we've posited. If our laws make ineliminable reference to a particular feature of the physical world and that feature is rather abstract, the problem may be even harder. For instance, we may be able to explain the failure of laws to be parity reversal invariant by following a suggestion from *Pooley* (2003) that there is a field that determines the parity or "handedness" of every physical process. The laws fail to be parity reversal invariant because they make reference to this parity field, so we have a fine explanation for why the laws are not parity reversal invariant but must still ask the further question of how we should understand this parity field. If, as an analogue of the third premise of the Argument for Temporal Asymmetry in Spacetime suggests, we should understand this structure as an element of spacetime structure, then we have a parity analogue for the Argument for Temporal Asymmetry in Spacetime. But we may resist this move and claim instead that we should posit another of Pooley's candidates to break the parity symmetry, namely a law that claims that all processes of a certain type have

---

[2]Such a structure must, of course, be connected to the temporally irreversible laws in such a way that this structure can explain the laws' temporal irreversibility.

the same orientation. Such as law is not best understood as an element of spacetime structure, and so we are faced with the question of both what object to posit to break the symmetry and how to interpret that particular object. Clearly, then, if the Argument for Temporal Asymmetry in Spacetime is to move forward, we need an argument in favor of premise 3 that tells us why we should interpret the structure posited as a result of premise 2 as an element of spacetime if we want to make the inference from dynamical asymmetries to asymmeries of spacetime. I examine the third premise of the Argument for Temporal Asymmetry in Spacetime in more detail in the fifth section of this chapter. In what follows, I will delve more deeply into each of the argument's premises to see exactly how well justified we are in believing the argument's conclusion.

6.2 An Extremely Brief Detour Through CPT

The Argument for Temporal Asymmetry in Spacetime's first premise tells us that not all of the laws of fundamental physics are TRI, but is this really so? Maudlin's passage above points us in the direction of the answer. Relativistic quantum field theories are typically taken to be invariant under the CPT symmetry, which maps quantum states to their charge-conjugated, parity-reversed, and time-reversed counterparts.[3] One of the important consequences of the CPT-invariance of quantum field theories is that these three different symmetries of charge conjugation, parity reversal, and time reversal are intertwined so that we can infer, from the failure of some quantum field theory to be CP-invariant, that the theory likewise fails to be T-invariant. The CPT symmetry thus allows for indirect measurements of T-violation. As it happens, there are at least two important cases of CP-violation, both of which involve the weak interaction. Both K meson decay and, more recently, B meson de-

---

[3]See *Streater and Wightman* (1980) for a rigorous derivation of the CPT symmetry for certain classes of relativistic quantum field theories.

cay have been found to violate CP symmetry, though I will not discuss the details of these experimental results here.[4]

What is important about CP-violating processes for our purposes, however, is that one can reason from experimental results involving K and B mesons to the first premise of the Argument for Temporal Asymmetry in Spacetime:

1. If the laws of a quantum field theory fail to be invariant under CP, then they fail to be invariant under T. (CPT theorem)

2. The laws of the quantum field theory governing the weak interaction fail to be invariant under CP. (Experimental results for the K and B mesons)

3. If laws of a quantum field theory fail to be invariant under T, then they aren't TRI.

4. Quantum field theories of the weak interaction are some of our best candidates for fundamental physical theories.

5. Not all of our best candidates for fundamental physical laws are TRI.
(from 1-4)

The argument is certainly valid, and premises 1, 2, and 4 seem well-supported by our current experimental and theoretical results in quantum field theory (though, of course, should we discover violations of CPT-invariance in quantum processes, then premise 1 should be rejected or explicitly limited to some subset of quantum field theories, perhaps not including the quantum field theory of the weak interaction, which would make the above argument unsound). Premise 3 is worth taking a closer look at since, as I argued in the first five chapters of this dissertation, TRI may

---

[4]See *Sachs* (1987), chapter 9 for a detailed discussion of K meson decay and *Abe et al.* (2001) for a discussion of B meson decay.

require something more (or maybe something less) than invariance under the operator T commonly appealed to in quantum field theories today. However, I will set aside a more in-depth discussion of this particular premise and simply remark that it seems a sensible premise to accept at this point in time, especially given the defeasible results at the end of the previous chapter. So, it seems like CP-violation in particle decays gives us a good reason to think that not all of our fundamental physical theories are TRI. So far, so good for the Argument for Temporal Asymmetry in Spacetime.

6.3 From Dynamical Asymmetries to Spacetime Asymmetries

There are many ways that one could justify the second premise of the Argument for Temporal Asymmetry in Spacetime, but it's worth mentioning the following principle in particular:

(SP2) All spacetime symmetries of a theory T are dynamical symmetries of T.

This symmetry principle is proposed and argued for by *Earman* (1989), and its contrapositive provides an important link between claims about temporally irreversible laws and fundamental temporal spacetime asymmetries. In fact, we can use SP2 to support premises 2 and 3 in the Argument for Temporal Asymmetry in Spacetime: if we have a dynamical asymmetry (an asymmetry in the laws), then SP2 tells us that we should have a spacetime asymmetry (presumably an asymmetry in an object representing a structure of spacetime) as well. As we'll see, however, SP2 only provides the Argument for Temporal Asymmetry in Spacetime with half the support it needs.

Before proceeding, it is important to start from Earman's own definitions of dynamical and spacetime symmetries to better understand what SP2 says and why it

might be true. The benefits of this analysis are twofold: first, we see how SP2 supports the second premise of the Argument for Temporal Asymmetry in Spacetime, and second, my analysis reveals a problematic move that prevents SP2 from supporting the third premise of the Argument for Temporal Asymmetry in Spacetime. Earman sets up the framework for his account of symmetries as follows: Let T be a theory whose models are of the form $\langle M, A_1, A_2, ..., P_1, P_2, ...\rangle$, where M is a differentiable manifold; A_i are the "absolute objects", geometric-object fields that characterize the structure of spacetime (more on these later); and P_i, the "dynamical objects", are the geometric-object fields that characterize the physical contents of the manifold.[5] For instance, a physical theory in Aristotelian spacetime would have models with absolute objects like a field picking out a preferred spatial origin and with dynamical objects such as fields picking out the spatiotemporal locations of physical objects, electromagnetic fields, etc.

Earman then defines spacetime and dynamical symmetries as follows:

Any diffeomorphism $\Phi : M \to M$ such that $\Phi * A_i = A_i$ for all i is a *spacetime symmetry of the model* $\langle M, A_1, A_2, ..., P_1, P_2, ...\rangle$. A diffeomorphism Φ is a *spacetime symmetry of a theory T* just in case it is a spacetime symmetry of all of T's models.

A diffeomorphism $\Phi : M \to M$ is a *dynamical symmetry of a theory T* iff $\Phi * A_i = A_i$ and $\Phi * P_i = P_i'$ for all i, where $\langle M, A_1, A_2, ..., P_1, P_2, ...\rangle$ and $\langle M, A_1, A_2, ..., P_1', P_2', ...\rangle$ are both models of a theory T.

A spacetime symmetry, then, is just some spacetime diffeomorphism that leaves the absolute objects unaffected, and a dynamical symmetry is a spacetime symmetry that transforms the dynamical objects in such a way that the result of the transformation still obeys the fundamental dynamical laws. In Earman's terms, then, the move

---

[5]More will be said about the distinction between absolute and dynamical objects in section 6.

needed to support the second and third premises of the Argument for Temporal Asymmetry in Spacetime goes as follows: we rely on SP2 to tell us that, if we have a dynamical asymmetry, then we have an asymmetry in the absolute objects. This step provides us with the support we need for the argument's second premise. But any asymmetry in the absolute objects is best understood as an asymmetric feature of spacetime itself, so SP2 allows us to take the third premise for granted and skip right to the conclusion of the Argument for Temporal Asymmetry in Spacetime. For the time being, we'll focus on SP2 alone and find that, while we may have a few reservations about this symmetry principle, overall SP2 (and, what's more, the second premise of the Argument for Temporal Asymmetry in Spacetime) is a reasonable symmetry principle to accept.

Earman gives two arguments in support of his claim that every spacetime symmetry of a theory is a dynamical symmetry of that theory, the second (and more compelling) of which is, essentially, that should we abandon SP2, we must also accept the unintuitive consequence that the laws of nature are not universal.[6] Imagine that we take some model $\langle M, A_1, A_2, ..., P_1, P_2, ... \rangle$ and apply the following transformation: leave the manifold M and the absolute objects A_i unchanged, but apply the time-reversal operator to the dynamical objects P_i. So, for instance, if we are dealing with a world governed only by the law of radioactive decay, we may find ourselves with a field on spacetime telling us how many particles of a radioactive substance we have at a given time. This field, for simplicity's sake, will have three values: 10 at time t, 5 at time $t+1$, and 2 at time $t+2$. Its time-reversed counterpart is, intuitively, a field with a value of 2 at t, 5 at $t+1$, and 10 at $t+2$.[7] Now, let's assume that the

---

[6]Earman's first argument contends that SP2 follows from the requirement that all the laws be generally covariant. Earman suggests that this argument for SP2 may be less than compelling since the link established by general covariance may introduce absolute objects that rule out dynamical symmetries of interest from theories that satisfy SP2.

[7]We do not need to take this transformation to give us the "real" time reversal operator since *Malament* (2004) and *North* (2008) have given us good reasons to think that the actual time reversal operator (if there is such a thing) does or could act on features of spacetime rather than just dynamical features like our "time reversal" operator here does.

laws are temporally irreversible but that time reversal is still a spacetime symmetry. Then the region in the original model and its time-reversed counterpart differ from one another only by a spacetime symmetry (time reversal), but we have a model (our original model) whose time-reversed counterpart isn't physically possible. This, Earman claims, violates our notion of universality, which he takes to be expressed by the following principle.[8]

> (U'): A putative law L is universal just in case, for all spacetimes governed by L, if R_1 and R_2 are two spacetime regions related to one another by a spacetime symmetry, then whatever L allows in R_1, L allows in R_2.

Thus, we have an inconsistent triad on our hands: we cannot simultaneously 1) accept that all laws are universal in the sense given by U', 2) take time reversal to be a spacetime symmetry of some theory, and 3) deny that time reversal is a dynamical symmetry of that theory. Earman suggests that, to rid ourselves of this tension, we reject either 2 or 3, and in doing so, we find that all spacetime symmetries are dynamical symmetries. To argue for SP2, then, one need only justify the claim that "all laws are universal".

We can understand the claim "all laws are universal" in at least two senses. In the first case, we can take this claim to be a necessary truth. I will call this position the "strong stance" since it suggests that being universal is partially constitutive of being a law. The "weak stance", on the other hand, suggests not that all laws have to be universal but only that universality is a theoretical virtue (like simplicity, perhaps). The weak stance suggests that, all other things being equal, if we are given the choice between two sets of laws, one of which contains only universal laws and one of which contains non-universal laws, we have reason to accept the first set of laws over the second set of laws. There may be non-universal laws, the weak stance admits, but

---

[8]Earman borrows this account of universality from *Rynasiewicz* (1986).

these laws will tend to be inferior to universal laws and may, for instance, provide less predictive and explanatory power than we would like our laws to provide.

Adopting the strong stance seems potentially problematic for a few reasons I'll list briefly.[9] Consider, for instance, Aristotelian physics, where one might posit a law like "All matter of such-and-such distance away from the earth is composed of quintessence". Such a statement leaves out a region of space and time (namely, the region occupied by Earth and bodies close to it), and so it fails to be universal. Another problematic case comes from *Tooley* (1977), who considers the putative law "All the fruit in Smith's garden are apples", which we may take to be confirmed when oranges brought into Smith's garden transform into elephants and bananas brought into the garden transform into apples. Laws like Tooley's, though fanciful, do seem to be metaphysically possible natural laws, and we might even accept their truth if we observed such strange instances of transforming fruit.[10]

Note also that many other kinds of "laws" can do the work we require of laws without needing to be universal. For instance, the rules of chess do a fine job of supporting counterfactuals, encoding regularities about how the game is played, and explaining various chess-related phenomena.; that is, the laws of chess govern the playing of chess in a very similar way to how the laws of physics govern the unfolding of events in our physical universe. But it would be absurd to think that, by introducing non-universal rules to our variant of chess, these new non-universal rules would now no longer be capable of playing the same role as the other rules of chess. If I now allow one special square on the chessboard to be immune to conquest by any other piece, we can posit a rule "No player may take a piece on E5", and we can rely on this rule to support counterfactuals like "If I moved my king to E5, your pawn would be unable to capture him", encode the regularity that no piece on E5 is ever captured,

---

[9]A more detailed catalogue may be found in *Lange* (1995).

[10]In Section 5, I'll show that, if we add a preferred spatial region as part of our spacetime structure, these particular laws may now qualify as universal in the sense of U'. However, unless we assume such additional structure, these laws fail to be universal.

explain the fact that no piece on E5 is ever captured, etc. So this new, non-universal rule seems in no way deficient when faced with the tasks one usually assigns to rules of chess, and if we are to believe the claims of the strong stance, we need a reason to think either that this analogy between the rules of chess and laws of nature doesn't hold or that there is some important function of laws that non-universal laws are unable to carry out. Examples like the ones I've just given provide us with a reason to think that U' may not be a constraint on lawhood as the strong stance suggests.

But even if universality is not a necessary feature of lawhood, it may still be a desirable one. Compare the rule "No player may take a piece on E5", which I shall call R1, with the rule "Bishops may move diagonally for as many squares as they like in any direction", which I shall call R2. R1 fails to be universal in that it singles out a particular region of the board as being special while R2 is universal in that it applies to all bishops, regardless of where they sit on the chessboard. R1 supports counterfactuals like "If I moved my king to E5, he would not be captured", but it is unable to support counterfactuals about what would happen outside of E5. R2, by contrast, supports counterfactuals about the behavior of bishops everywhere on the chessboard. R1 encodes regularities about what happens on E5 but not about what happens elsewhere while R2 encodes regularities about the behavior of all bishops everywhere on the chess board. R1 may be invoked to explain a limited range of facts about what happens (or doesn't) to pieces on E5 while R2 is capable of explaining a wider range of chess-related phenomena we might observe. So while my chess example may suggest that the non-universal rules are capable of doing the same kind of work that universal rules can do, it seems that universal rules are generally better at performing the tasks usually assigned to rules than non-universal rules are.

If the analogy between laws of nature and the rules of chess holds here, then, we have a good reason to adopt the weak stance towards SP2. Universal laws may not be the only laws available to us, but they may be the best sort of laws we can

rely on to do the sort of work we require laws for. Adopting the weak stance does require us to change our view of what the Argument for Temporal Asymmetry in Spacetime establishes a bit: the reason why we should posit a temporally anisotropic spacetime given temporally irreversible laws is that we believe physical theories that posit temporally irreversible laws and temporally anisotropic spacetimes have at least some theoretical virtues that physical theories that posit temporally irreversible laws and temporally isotropic spacetimes lack. The conclusion of the argument is softened since there may be other theoretical considerations (e.g. simplicity) that also help us decide what to believe and may, hypothetically, lead us to reject a theory with only universal laws in favor of a theory with some non-universal laws.

So the weak stance towards SP2 gives us a reason to believe the second premise of the Argument for Temporal Asymmetry in Spacetime. There are, however, simpler arguments for those who dislike SP2 but still want to argue for the argument's second premise. When we are faced with dynamical laws that fail to be TRI, we may want to explain this fact, especially in a case like K meson decay where we have a temporally irreversible law that we discover after a long string of other physical laws that are all TRI. "What makes K meson decay special?", we may wonder. We are left with a few different options for how to explain a law's failure to be TRI. One option is to posit that that's just the way the law is: laws, we may think, are brute facts about our physical world, so we cannot provide a deeper explanation for their features. But there's an alternative: we can explain a feature of physical laws by referring to some other structure (in our case, a temporally asymmetric structure) on which the laws depend. Such a structure improves our theory because we now have an explanation for a fact (the TRI-failure of some of the laws) that we didn't have before. Theories that explain more phenomena are better than those that explain a more limited range of phenomena, so we may have a reason to believe the argument's second premise even if we don't believe universality as defined by U' gives us a compelling reason to adopt

the weak stance towards SP2.

But we must be careful with the object that we posit to break the TRI of the laws we are concerned with. As I pointed out above, theoretical virtue considerations typically come with a trade-off: by choosing a theory that explains the TRI-failure of the laws instead of treating it as a brute fact, we're positing the existence of some new object that the "brute fact" theory didn't posit. If we're going to harm virtues like parsimony in our search for explanatorily satisfactory physical theories, we should do as little violence to these virtues as possible. What we're looking for, then, is a minimal structure that we can rely on to break the TRI of our fundamental laws, at least if we want our motivation for accepting premise 2 to lead us to accept premise 3 of the Argument for Temporal Asymmetry in Spacetime as well.

Let's set aside the details of that minimal structure for the moment, though, and consider how far Earman's SP2 can get us. If we have a way of showing that spacetime asymmetries, in Earman's sense, should be interpreted as being asymmetries of spacetime itself, then we have an argument in favor of the Argument for Temporal Asymmetry in Spacetime's third premise that we need and we need not consider the issue further. Let's proceed with some caution here: according to Earman's definition, if we posit some object that breaks a spacetime asymmetry, then that object must be an absolute object that isn't left unchanged by the transformation in question. This follows from the fact that, by definition, a spacetime asymmetry is going to reveal that at least one absolute object isn't going to be left unchanged by the transformation in question, so positing dynamical objects alone won't help us explain this *spacetime* asymmetry. What we need to show, in order to establish SP2's support for premise 3, is that Earman's absolute objects need to be interpreted as features of spacetime.

This may not seem too hard a thing to prove. After all, *Earman* (1989) takes absolute objects to represent the "fixed space-time structure" (45). Earman does

not provide us with a formal definition for his absolute objects, but he endorses the account shared by other authors such as *Friedman* (1983), who does provide such a formal definition. Friedman's definition of absolute objects is essentially as follows:

(A) A geometric object Φ_i is an *absolute object* of a theory T just in case for any two models $\langle M, \Phi_1, ..., \Phi_n \rangle$ and $\langle M, \Psi_1, ..., \Psi_n \rangle$ of T, for every $p \in M$, there are neighborhoods A, B of p and a transformation $h : A \rightarrow B$, such that $\Psi_i = h\Phi_i$ on $A \cap B$.

What is clear from this definition is that being an absolute object is neither necessary nor sufficient for being interpreted as a feature of spacetime itself. For instance, take any non-vanishing vector field, such as the Newtonian gravitational field or the electromagnetic field. It follows from the vector straightening theorem and Friedman's definition together that any such vector field constitutes an absolute object, yet we clearly would not want to interpret any such object as representing something about the structure of spacetime. A similar result can be established for any constant scalar field. Friedman also points out that, while the metric tensor is an absolute object in special relativity, it is not so in general relativity; since the metric tensor provides us with important information about the structure of spacetime in general relativistic theories, it then seems clear that being an absolute object also isn't necessary for representing the structure of spacetime. The question of whether or not the mathematical object that breaks the TRI of the dynamical laws is an absolute object of our physical theories is thus entirely separate from the question of how we should interpret that object. SP2, then, may give us a reason to accept to accept the second but not the third premise of the Argument for Temporal Asymmetry in Spacetime.[11]

---

[11]Earman could, perhaps, take an analogue of the weak stance towards premise 3; that is, one could argue that, while not all absolute objects represent spacetime structure and not all elements of spacetime structure are represented by absolute objects, it's usually best to prefer theories that require that all and only absolute objects represent spacetime structure to theories without such a requirement. Such a stance seems especially reasonable given that the context of Earman's positing SP2 is a lengthy discussion of classical spacetime theories, for which the weak stance towards inter-

6.4 An Object to Break the Symmetry

If SP2 cannot support the third premise of the Argument for Temporal Asymmetry in Spacetime, we'll have to look elsewhere, but first, we should return to the question of what a reasonable temporally asymmetric structure to posit might be. One option that should immediately occur to us is Price's "Machian" law. As an analogue of Pooley's parity law, we could posit a law that says "all sequences of events have the same temporal orientation" and add this law to our set of fundamental laws. We then need only stipulate at some point in spacetime which direction is "past" and which direction is "future" to fix the temporal orientation at all times. Such a law governs features of sequences in spacetime but, like other laws like those of statistical mechanics, should not be interpreted as a feature of spacetime, so it provides us with sufficient grounds to reject the third premise of the Argument for Temporal Asymmetry in Spacetime. There are those who might object to this law on the grounds that it is non-local because a stipulated orientation in one region of spacetime determines the temporal orientation at all points in spacetime, so, in fairness to the proponent of the Argument for Temporal Asymmetry in Spacetime, I'll consider a different object.

The following structure also serves as a temporally asymmetric physical object that we could introduce to our ontology to account for temporally irreversible laws. Take the orientation field ϕ to be a map from spacetime points to light cones; essentially ϕ allows us, at each point in spacetime, to pick out a single unique, consistent light cone which we take to be the future light cone at p.[12][13]. There are two different

---

preting absolute objects as features of spacetime is quite well-motivated. An advocate of this weak stance will still need to account for my examples and explain why the weak relationship between absolute objects and spacetime structure is acceptably violated in cases like general relativity, but it should not be dismissed out of hand.

[12]The notion of consistency I have in mind here comes from *Earman* (1974). The basic idea is that there is no point x in spacetime such that continuous transport of the light cone around a closed loop based at x causes points previously on one side of the light cone to now fall in the other side of the light cone.

[13]As the earlier quote from *Price* (2011) suggests, though, the field ϕ' that picks out the past light

interpretations of ϕ that will be important for our purposes. The first, which corresponds, roughly, to Horwich's and Maudlin's proposal, takes the object ϕ to be a feature of spacetime; that is, the first interpretation takes ϕ to mathematically represent geometric features of spacetime itself. Depending on whether we're relationists or substantivalists about spacetime, we may understand ϕ differently, but I will lump all such interpretations under the same heading for my purposes here. The second interpretation, on the other hand, takes ϕ to be a mathematical description of a feature of matter fields, like a temporal arrow with which certain kinds of matter come equipped, much like the electric field in classical electromagnetism. Put differently, this second interpretation takes ϕ to be a map from points in spacetime to structural features of spacetime even though ϕ itself is not an element of the spacetime structure. One may draw parallels between this interpretation of ϕ and relationist theories of spacetime, but I see no reason to think that only relationists can understand ϕ this way. Our first interpretation takes ϕ to encode geometric features of spacetime while our second interpretation takes ϕ to encode features of the contents of spacetime. I will call the first interpretation "geometrical" and the second interpretation "material", and I will refer to the kind of object each takes ϕ to be by using the name of these interpretations.

The difference between these two interpretations may be clearer if we consider the following example: imagine that we have discovered that all particles of a certain type rotate clockwise about an axis pointed in the same direction z. We can represent this fact mathematically by positing a field ξ which is a map from points in spacetime to a spatial vector in spacetime representing the direction of the particle's rotation using the right hand rule. We can interpret this field ξ in two ways: if ξ is a geometrical object, then we treat this vector field as representing facts about the structure of spacetime. There is a preferred direction in spacetime, and it's the direction in which

cone at each point in spacetime would work just as well for our purposes.

146

all of these particles are pointing. If ξ is interpreted as a material object, however, then we treat this vector field as representing a feature of the matter fields of those special spinning particles. We are then faced with the same problem concerning ξ that we now face regarding ϕ: which of these two interpretations gives us the best way of understanding our new mathematical object? In other words, should we think of a temporal orientation field like the metric tensor or like an electric field?

With the background set and our options laid out, we can now consider arguments for or against these positions which, in turn, will allow us to support or refute the third premise of the Argument for Temporal Asymmetry in Spacetime. One first intuition might be that Occam's razor can tell us whether we should believe a particular mathematical object to be material or geometric since, after all, we must balance the desire for a wide explanatory scope that led us to posit ϕ in the first place with other theoretical virtues like parsimony. But this intuition leads nowhere since we're still dealing with the same number of things whether we add a material or a geometrical object into our ontology as a consequences of the TRI-failure of the fundamental laws. Appeals to many other theoretical virtues like ease of Popperian falsifiability will not help us. Treating a particular mathematical object as geometrical will not make our physical theory any easier or harder to falsify since these two options are empirically indistinguishable. It seems that the philosopher of science's usual bag of tricks may be of limited use here.

But there is one notion from the general philosophy of science, namely methodological conservatism, that can do some real work for us. It may not settle the issue, but it will weigh in favor of the materialist's position. The conservative principle best suited to this project is the one that licenses the inference in the following passage from *Sklar* (1975):

We have to choose among alternative hypotheses all of which seem equally warranted on the basis of reasonable inference from the evidential data.

147

But suppose one of the hypotheses is "more in conformity with preexisting theory" – more "like it in structure," say, or more "conceptually continuous" with it. Then that hypothesis is preferred. (31)

In other words, methodological conservatism tells us to prefer the hypothesis that deviates as little from our previously-held hypotheses as possible. We can apply this principle to the geometrical/material debate as follows: though we may not have a clear account of how to determine which objects are geometrical and which are material, most authors agree about many of the features that we should interpret as geometrical according to particular physical theories. The metric tensor and the properties we can determine from it such as spacetime curvature, the connection, light cone structures, and the spacetime separation are always taken to be geometrical, not dynamic, objects, at least in our best currently available theories of space and time. Thus, the methodologically conservative among us will want our new geometrical objects to stick as close to these original geometrical objects as possible. Let us then provisionally state that we should take the geometrical objects to be the ones that we can derive from the metric tensor alone. We may posit new sorts of geometrical objects if empirical data or other theoretical virtues drive us to do so, but, all other things being equal, it's better to stick with theories that take as geometrical only the sorts of objects we're used to taking as geometrical.[14]

But why should we be so resistant to changes in geometrical objects? Our best theories of spacetime physics are our best theories for a reason, namely because they are empirically successful and have been for a while. Those of us who are scientific realists typically take the empirical success of a scientific theory to provide a (po-

---

[14]One virtue of this conservative principle that's worth remarking on is that it lines up quite nicely with traditional philosophical accounts like Quinean webs of belief and Lakatosian research programmes. Both of these philosophical accounts appeal to a network of relations among the commitments of our theories, and we clearly take some of these commitments to be more central or important than others. When our theories confront empirical evidence and something must be rejected, we should begin by rejecting the less central claims before we reject any truths more central to our theory. And this is precisely what my conservative principle recommends we do.

tentially defeasible) reason to believe that the scientific theory has gotten certain features of the world right. We may no longer believe that Newtonian mechanics is literally true, but we do believe that it got a number of important things right about the relationship among, for instance, the mass, velocity, acceleration, and position of medium-sized physical objects. Likewise, we may think that our current set of geometrical objects posited by our best available physical theories is incomplete because, like all physical theories, it will likely be modified or discarded in the future; however, since we are not in a position to know what it is that our current set of geometrical objects gets right, the best policy would be to stick to this set as closely as possible while still allowing these objects to play the role in our physical theories that they currently play.

Having identified the metric and what follows from it as our litmus test for identifying geometrical objects, we can now apply our principle of methodological conservatism to help us determine when we should identify a particular object as geometrical or material. The metric tensor is a function defined on a spacetime manifold that maps a pair of vectors tangent to that manifold to a scalar representing the spacetime separation between these vectors. The metric tensor is symmetric in both space and time; that is, for any metric tensor g_{ab}, it follows that $g_{ab}(\mathbf{v}, \mathbf{w}) = g_{ab}(\mathbf{w}, \mathbf{v})$ for any vectors $\mathbf{v}, \mathbf{w}$ in the tangent space. So the metric alone does not tell us whether, for instance, a point p which is time-like separated from a point q is in q's future or q's past any more than it could tell us whether the point r which is space-like separated from a point q is to the left or right of q. If the metric alone does not give us this sort of information, we need to rely on some sort of supplemental structure to give it to us. We have two options: we can treat this asymmetrical structure as a geometrical or material object. If we treat it as a geometrical object, however, we risk violating my conservative principle. After all, relativistic metric tensors do not provide us with any information about preferred directions in space or time; why should we expect

the metric tensor found in our new theory to play such a different role?

So methodological conservatism may give us reason to think that any orientation field like the one I've proposed is best treated as a material rather than a geometrical object because it performs a task separate from the kinds of tasks performed by the metric tensor in the recent past. However, this reason is not decisive. Other considerations may lead us to believe, for methodologically conservative reasons, that interpreting ϕ materially is incorrect since it has many features (not carry energy, providing only a direction but not a magnitude, etc.) that many other material fields have. When weighing such cases against one another, it still seems clear to me that interpreting ϕ materially is the *more* methodologically conservative of these two options, but, to keep the discussion from digressing into table-thumping about intuitions, we may have to accept the fact that methodological conservatism alone may not be sufficient to resolve this issue. But this result is perhaps unsurprising; had we followed methodological conservatism alone, we would have refrained from discarding Newtonian physics in favor of special relativity. The suggestion that every new mathematical object should be treated as material instead of geometrical seems absurd, and one might worry that my methodology could saddle us with the view that the compensatory Lorentz theory is preferable to special relativity, which is a disagreeable consequence to say the least. What is needed, then, is some way to distinguish the cases where we do want to posit new geometrical objects like special relativity from cases where the new symmetry-breaking objects seem best interpreted as material. This reason, hopefully, will give us a way out of the methodologically conservative gridlock.

In this situation, we're forced to take a stance on just what it is we want spacetime to do for us in our physical and metaphysical theories. Consider the case of special relativity. What was it that motivated physicists to posit the existence of spacetime as opposed to simply remarking on the fact that all of the laws seemed to be Lorentz-

invariant? The value of spacetime is that it provides a unifying explanation for a diverse range of phenomena that we may think cry out for explanation. Spacetime here serves as a kind of meta-law, explaining and justifying the Lorentz-invariance of the physical laws instead of requiring that we take the Lorentz-invariance of these laws as a brute fact. So, if all the laws seem to be Lorentz-invariant, then we have a fact about these laws to explain, and geometrical objects may be posited to explain this fact.

But imagine that we are given a case like Tooley's garden where we find that one particular region of spacetime is "special" in some sense. Perhaps, for instance, we have a physical theory that fails to be spatial translation invariant (STI) because the laws allow some things in a region X to be possible that are not possible outside of X. Should we posit that the region privileged by this strange law is somehow privileged by the spacetime structure of our world? Well, it depends on what facts we need to explain. If all of our other physical theories are STI, then it seems the only fact we have to explain is the failure of one particular natural law to be invariant under STI. We may explain this by tracing the STI-failure back to initial conditions that fail to be STI despite STI micro-laws, or we may posit some object that fails to be STI that the laws couple to. But to posit that this preferred region X is a special geometrical object is to leave more unexplained than it ever explained in the first place. If spacetime itself fails to be STI, then why are so many of our other laws STI? If we were to investigate and find that, in every case, STI-failing laws were more empirically adequate than their STI alternatives, then we wouldn't be faced with this explanatory embarrassment by positing some spacetime structure that failed to be STI.[15] So it seems like the explanatory gains outweigh the costs when we posit an asymmetry in spacetime that every dynamical law seems to exhibit; however, if most of the laws are invariant under some symmetry and we have only one or two laws

[15]Note that, by incorporating a special position as an element of spacetime structure, we have now made the law in the Tooley's garden example universal in the sense of U'.

that are not, then positing an asymmetric feature of spacetime to explain these few asymmetries comes at a great explanatory cost, and the benefits just aren't worth it.

Put differently, I'm suggesting that we adopt the following rule, which is just a variant of inference to the best explanation: we have more reason to assign interpretation I to mathematical object o than we have reason to assign interpretation I' to o only if I provides us with a better explanation for the physical phenomena that motivated our positing o in the first place than I' provides. There may be difficult cases where reasonable individuals disagree over which theory or interpretation is truly more explanatory, but in many situations, our paradigm cases of excellent explanations (e.g. those that are more unifying, those that are less ad hoc, etc.) will provide us with an acceptable metric by which to judge which theory or interpretation is more explanatory and will frequently lead to general agreement. My rule does not tell us that the more explanatory theory or interpretation is true, nor does it suggest that we not continue investigating theories or interpretations that we do not believe at this time. It does suggest, however, that the currently available evidence should lead us to believe the theory or interpretation that provides us with better explanations and fewer unexplained phenomena than its rivals.

For our case, that means that we should only interpret ϕ as geometrical if all or most of the laws fail to be TRI. If most of the laws are TRI, then a geometrical ϕ would need to couple to some laws but not others, and we would need some sort of explanation for this feature of spacetime that, unlike every other feature of spacetime we've posited to date, only affects certain laws but not others. A material ϕ, on the other hand, would not have so difficult a burden as it could be interpreted as merely describing geometric features of the matter governed by the temporally irreversible laws. As I discussed previously, the CP-violations that lead us to believe that the fundamental laws of physics are not all TRI have only been observed in two different types of mesons. We have no reason, at present, to believe that the symmetry-

breaking is more widespread than that since CP-violations have not been observed in experiments involving numerous other particle collisions. Even if we take the standard model alone as our best available physical theory, the vast majority of the terms in the Lagrangian we would write down to describe all of the particles in the universe would be TRI. If we posit, then, that there is a particular field, ϕ, that some mesons couple to and that the rest of the particles don't, we have an explanation (or at least the beginning of an explanation) for the TRI-failure we observe. If we posit that spacetime itself fails to be TRI, however, we must explain why the Lorentz-invariance of spacetime makes every term in the Lagrangian Lorentz-invariant while the TRI-failure of spacetime doesn't actually imply that most of the terms in the Lagrangian for the universe fail to be TRI.

For those who remain unconvinced, let's consider an analogy with the electric field. When faced with phenomena suggesting the presence of an electric field, physicists could have chosen to interpret this field geometrically or materially, but in the end they chose to interpret it materially. What could their reasoning have been? At least part of it, I would think, is the fact that not all material objects interact with the electric field in interesting ways. Uncharged particles would not have their motion directly affected by such a field, and so it doesn't seem suitably universal to be interpreted geometrically. This example may lead to a new interpretive principle that even those wary of my explanatory principle would be willing to adopt: geometrical objects have a wider (or more universal) scope than material objects, so an object should only be interpreted geometrically if its scope is sufficiently wide. Considering the salient universality standards used to apply this principle to the case of the electric field, we find that the electric field should be interpreted materially, and it seems hard to see how such standards ruling the electric field a material object could consistently support interpreting ϕ as a geometric object. Thus, considerations of other materially-interpreted fields suggest that, due to considerations of scope, it's better to interpret

ϕ materially than geometrically.[16]

Treating ϕ as a material object leaves us with less to explain than treating it as a geometrical object does and better aligns with our treatment of fields like the electric field in the past, so we have a good reason to think that ϕ is a material rather than geometrical object; however, should we find that other physical theories are improved by setting them in spacetimes containing a geometrical object like ϕ, then we may need to revise this opinion. Without many such physical theories to appeal to, we seem to have a good reason to reject the third premise of the Argument for Temporal Asymmetry in Spacetime by taking ϕ to be material, and so we have a good reason to reject the conclusion of the Argument for Temporal Asymmetry in Spacetime.

6.5   What is Spacetime and How Should We Treat it?

At this point, it's worth considering an objection to my diagnosis of when we should treat a mathematical object as geometrical that arises from my characterization of spacetime. There is a fairly intuitive view that I have endorsed here which can be found in works by *Balashov and Janssen* (2003) and *Janssen* (2009) as well, namely that the spacetime of a physical theory is more fundamental than the dynamical laws and provides the spatiotemporal relations that the laws rely on to help us predict and explain physical phenomena. Consider, for instance, any inverse square law such as Coulomb's law or Newton's law of universal gravitation. These laws allow us to determine the motion of particles under the influence of certain kinds of forces, and the influence of the force is proportional to the inverse square of the distance between the particle and the origin of the force. It is the fundamental spacetime relations that tell us the distance between the particle and the force's origin. The dynamical laws rely (both explanatorily and ontologically) on spacetime in that facts

---

[16]I believe that arguments similar to mine would support a material rather than geometrical interpretation of ϕ's spatial counterpart posited to explain failures of parity reversal invariance, but the implications of my approach to the topic of parity are a topic for another work.

about distances, for instance, will determine the form certain laws take. The basic picture is of an underlying, unifying spacetime that fundamentally grounds facts about the dynamical laws and allows us to explain the behavior and symmetries of various laws by appealing to the spacetime of the theory from which these laws come.

There is, however, an alternative picture available, one which denies the move I make in my diagnosis and provides independent support for the second and third premises of the Argument for Temporal Asymmetry in Spacetime. *Brown* (2005) and *Brown and Pooley* (2006) suggest that Janssen has "put the cart before the horse" (*Brown* (2005), 132) in this case and that it is the laws of physics that ground, justify, and explain our claims about spacetime.[17] For instance, it is not the case that the laws of physics are Lorentz invariant because spacetime itself is Lorentz invariant; rather, we claim that spacetime is Lorentz invariant as a way of capturing an important fact about our dynamical laws, namely their Lorentz-invariance. Spacetime is, according to Brown and Pooley, a "glorious non-entity", completely incapable of explaining facts about the dynamical laws in the way that Janssen suggests. It seems fitting to think of Brown's account as taking a similar view towards spacetime as the advocate of Humean supervenience takes towards natural laws: just as the Humean claims that natural laws are no more than generalized regularities that satisfy some important cognitive criteria, Brown claims that spacetime is no more than a mathematical representation of certain features many of the laws seem to share.[18] It is the laws that explain and justify facts about this spacetime structure, not the other way around. For instance, *Brown* (2005) claims "...the Minkowskian metric is no more than a codification of the behaviour of rods and clocks, or equivalently, it is no more than the Kleinian geometry associated with the symmetry group of the quantum physics of the non-gravitational interactions in the theory of matter"

---

[17]The fact that Brown makes claims about ontological as well as explanatory priority here is noted by *Norton* (2008).

[18]A discussion of Brown's account on page 128 of *Frisch* (2011) makes these similarities quite striking.

(9). Relying on a causal account of explanation, Brown suggests that Minkowski spacetime does no real explanatory work: it can't cause matter to move in certain ways, so how can we take it to explain why matter moves as it does? Given the explanatory deficiencies of spacetime, Brown suggests we do the parsimonious thing and cut down on our metaphysical commitments by removing spacetime, insofar as it represents something separate and apart from a codification of facts about the physical laws, from our ontology.[19]

If Brown is correct, then not only is there no real, meaningful geometrical/material distinction here, but there is also a fairly straightforward argument for asymmetries in spacetime from dynamical asymmetries; after all, if there is nothing more to spacetime than what's already encoded in the facts about the dynamical laws, then finding out that the dynamical laws are temporally asymmetric should lead us to believe that the spacetime we posit on their account is likewise temporally asymmetric. If Janssen is correct, however, the connection between spacetime and dynamical asymmetries should be something subtler, perhaps along the lines of what I've suggested in the previous sections. In order to head off objections to the view I've laid out in the previous sections, then, I must provide compelling reasons to reject Brown's account in favor of Janssen's. Our first attempts to determine which of these two views is correct may rely on intuitions about physical possibilities. For instance, we may ask ourselves whether it seems physically possible for there to be some spacetime with more (or less) structure than required by the dynamical laws. An example: we could consider Lorentz-invariant dynamical laws in spacetimes that are not Lorentz-invariant. Are models such as these metaphysically possible? I see no reason to think they are not, but these intuition-pumping cases don't provide much in the way of argument for Janssen's position over Brown's or vice-versa if intuitions diverge

---

[19]To my knowledge, the first Brown-like view on the nature of spacetime was discussed in III.F of *Sklar* (1977).

widely in these cases.[20] More sophisticated arguments are clearly required here.

We turn, then, to issues of explanation, which is where Brown and Janssen themselves seem to take the best arguments in favor of their respective positions to lie. Brown argues that there is a kind of gap in explanations that rely on spacetime to explain, for instance, why moving rods contract, or, in the case of general relativity, why light rays follow geodesics. As he puts it, "The real issue is not whether physical geometry is easy to get your hands on, but rather whether, when it is absolute and immune to perturbation as in Newtonian and Minkowski space-time, it offers a causal explanation of anything" (*Brown* (2005), 26). The problem, according to Brown, is that spacetime cannot provide causal explanations of material phenomena, so it cannot explain them at all. If we want to explain why a law is Lorentz-invariant, for instance, Brown suggests that an adequate explanation would be to show how to derive this law from a more fundamental Lorentz-invariant law. If the law in question is fundamental, however, then Brown does not provide us with any resources to explain this fact. Regardless of the law's status, it would be unacceptable, Brown claims, to try to explain the Lorentz-invariance of a law by appealing to the structure of spacetime as an entity in its own right.

Janssen, for his part, agrees with Brown that, if spacetime is explanatorily inert, it would be better to treat spacetime as Brown's "glorious non-entity". However, he draws on work from *Lange* (2007) to argue that, without spacetime, there's a particularly odd fact that Brown and members of his camp need to account for. In particular, special relativity is a Lorentz-invariant theory, and this feature, Lorentz-invariance, is assumed to be true from the start in other theories such as the electroweak theory

---

[20]Earman's symmetry principle SP1, which takes every dynamical symmetry to be a spacetime symmetry, is supported by the claim that spacetimes with more structure than is required by the dynamical laws violate Ockham's razor and could be of use to the advocate of Brown's position here. But note that Earman's symmetry principle and his argument supporting it only establish that theory-interpretation pairs that don't violate Ockham's razor are better than those that do. Brown needs to establish something a bit stronger to respond to me here, namely that spacetimes that violate SP1 are, in some sense, conceptually incoherent or impossible.

and quantum chromodynamics. The laws of our best available physical theories all seem to invariant under the same set of symmetries, which, on the face of it, seems quite odd; in fact, Janssen takes this feature of laws that all of our best available physical laws of nature seem to share to call out for an explanation. Without some unifying explanation, it seems quite the coincidence that all of our best available laws just so happen to share this feature in common.[21]

There is a possible response available to Brown here: he can appeal to an underlying physical theory invariant under the symmetries that require explanation. For instance, Brown could argue that many of the laws we use to are Lorentz-invariant because the fundamental physical theory that underlies them is also Lorentz-invariant. This line of response is promising, but it still seems problematic for several reasons. First, what if there isn't a single dynamical physical theory that underwrites all of the laws in question? If this isn't so, we can then reformulate Janssen's worry by asking what explains the invariance of these more fundamental laws under the symmetries they're invariant under. Brown's possible line of response is better off if there is only a single physical theory that explains the symmetries of all of these disparate physical theories like the "grand unified theory of everything" (GUTE) discussed in chapter 2. If such a theory exists, brute facts about that theory seem no worse off explanatorily than brute facts about spacetime. The worry for this line of response, though, is that pluralism in physics may win the day and no empirically adequate GUTE may ever be found. Because the line of response considered here relies on the existence of a physical theory that's not currently available to us, I think it is reasonable to discount this response to Janssen until such a theory is discovered.[22] So, absent a GUTE, the

---

[21]This point is made on page 182 of *Frisch* (2011) as well.

[22]Some may worry about a tension here between my agnosticism towards the existence of a GUTE and my discussion of GUTEs in chapter 2. Note that, in chapter 2, my purpose was to show how a particular account of symmetries would be useful to physicists and philosophers in their respective projects. Searching for a GUTE is a major project of interest to physicists and philosophers alike, so the fact that my account of symmetries will be useful for those working on GUTEs is still a sign of its pragmatic value even if, ultimately, we can formulate no empirically adequate GUTE.

best line of argument available to Brown in response to Janssen is to argue that it's acceptable to take some facts about laws as brute.

Both camps can dig in their heels here: Brown can claim that the fact that all of the best fundamental laws seem to be Lorentz-invariant is not the sort of fact that calls out for explanation (or may even be the sort of fact that *can never* be explained) while Janssen can claim that the fact that moving rods contract is explained (insofar as it can be explained) by the nature of spacetime but that not all explanations need be causal explanations. The current state of the debate between these two views on spacetime seems to rest on the reasonableness of these two explanatory claims, and one may again worry that we've reached a point where intuitions about explanation are all we can rely on to settle the matter.

But this is not the case. Brown claims that all genuine explanations are causal, yet there seem to be good reasons to think that causal accounts of explanation provide us with too narrow a picture of what it takes for an explanation to be genuine. Causal accounts of explanation tell us that "A explains B" is true just in case there is a causal relationship between A and B such that A (or a substructure of A) causes B (or a substructure of B), but there are numerous counterexamples that seem to suggest otherwise. For instance, we may explain a drop in the pressure of an ideal gas by appealing to the gas's increasing volume, and we may explain an increase in an ideal gas's volume by appealing to a drop in the gas's pressure. Yet causal relationships are asymmetric since causes temporally precede their effects. So no matter what causal story we decide to tell about the ideal gas, one of our explanatory claims will be unaccounted for.[23] Causal accounts of explanation are similarly poor at accounting for explanations of laws that appeal only to other laws or meta-laws. Unless we're willing to accept the claim that features of one law can cause features of another law, Brown may have a hard time explaining why he takes it to be the case that the

---

[23]See *Price* (1997) for an account of backwards causation which one could use to dispute claims about temporally asymmetries in causation.

Lorentz-invariance of law L can be explained by the Lorentz-invariance of the more fundamental law L' from which L can be derived. The relationship here does seem to be explanatory, but, at least on most common accounts of causation, it doesn't seem to be distinctly *causal*. We have good reason, then, to believe that causation is not a necessary condition for explanation.

I bring these problems with the causal account to the fore because there's a better account of explanation available, namely one that relies not on the notion of causation but on the more general notion of dependence. *Jansson* (2011) (not to be confused with Janssen) has recently advocated an account of explanation that runs essentially as follows: we can truthfully say "A explains B" just in case both A and B occur and, when A occurs, B also occurs in virtue of A, and when A does not occur, B fails to occur in virtue of A's failure to occur. These dependency relationships that Jansson relies on are more general than causal relationships, so they seem better suited to deal with my previous counterexamples that showed that causation was not necessary for explanation. On Jansson's account, the ideal gas law encodes a nomic dependency relation between the volume and pressure of an ideal gas, so the pressure of the gas can explain its volume and vice-versa, even if there is no causal relation we can posit to justify both explanations. And features of the laws are certainly dependent on (if not caused by) features of more fundamental laws. We thus have good reason to prefer the dependency account of explanation to the causal account.

Armed with the dependency account of explanation, we may attempt to break the explanatory stalemate in the Brown/Janssen debate by seeing how well each account's claims about explanation fare when we take dependencies, not just causes, to underwrite genuine explanations. We now seem to have good reason to think that spacetime provides us with a genuine explanation for phenomena like contracting rods. These phenomena depend on the structure of spacetime since moving rods contract *in virtue of* the Lorentz-invariance of spacetime and a relation between the

contents of spacetime and spacetime relations that connect them to one another. If the structure of spacetime changes, we can expect this phenomenon to likewise change, so the sort of dependency relation Janssen might posit in defense of his view of spacetime seems to fit the explanatory criteria laid down by the dependency account of explanation. Brown may object that Janssen needs to say a bit more about the relations that obtain between spacetime and material objects, and this may be desirable for a more satisfying explanation of special relativistic phenomena, but such details aren't necessary for showing that spacetime explains phenomena in special relativity. Likewise, spacetime can explain the fact that laws in theories like quantum chromodynamics and the electroweak theory are Lorentz-invariant. The relation between spacetime and these theories is such that we can determine the Lorentz-invariance of these theories given only a Lorentz-invariant spacetime and the relation between spacetime and the dynamical laws. If the structure of spacetime changes so that it is no longer Lorentz-invariant, the relation between spacetime and the laws requires a change in the Lorentz-invariance of these laws as well. Again, we have a genuine explanation of a phenomenon that requires us to appeal to spacetime.

So where does this leave us? Brown has argued that his account of spacetime is preferable because it provides an explanation for facts that Janssen's account can't explain, namely facts about moving rods and clocks in special relativity. Janssen agrees that, were this true, it would indeed give us a reason to prefer Brown's account, but Brown's claim about the explanatory inadequacy of spacetime is false. As I have shown, Janssen's spacetime, when coupled with a dependency account of explanation, is capable of explaining not only the phenomena Brown mentioned but also a phenomenon that Brown's account must take as brute: the fact that all of the best physical laws seem to be Lorentz-invariant. Thus, since Janssen's spacetime is capable of doing all the explanatory work Brown claimed it could not while also explaining phenomena that Brown's account is unable to explain, we have a compelling

reason to favor Janssen's view of spacetime and a response to objectors who would rely on Brown's account to deny my geometrical/material distinction and justify the second and third premises of the Argument for Temporal Asymmetry in Spacetime. At the very least, my response has shown the work cut out for those who would support the Argument for Temporal Asymmetry in Spacetime by appealing to Brown's account: they must show why the causal account of explanation is superior to the more general dependency account of explanation, which seems a difficult task indeed.

6.6   Conclusion

To quickly recap, I have argued that, though several philosophers have proposed accounts that seem capable of licensing an inference from temporally irreversible laws to a temporal spacetime asymmetry, none of these accounts or arguments are up to the task. The Argument for Temporal Asymmetry in Spacetime seems promising, and we can support its premises by appealing Earman's symmetry principle SP2 or by appealing to explanatory considerations. While the argument's first premise is reasonable, the best support we can find for the second premise involves taking a weak stance towards SP2 and likewise weakening the argument's conclusion. More importantly, however, Earman's SP2 doesn't establish the existence of a symmetry-breaking object that is best interpreted as geometrical. I have argued that we have two ways of breaking the symmetry: we can posit a "Machian" dynamical law, which leads us to reject the third premise, or we can posit an object ϕ which, except in cases where *most or all* of the laws fail to be TRI, we have a good reason to think must be material instead of a geometrical regardless of our substantivalist/relationist sympathies. So, even if one accepts the second premise of the argument, the third premise is false, and the Argument for Temporal Asymmetry in Spacetime fails. What is more, while Brown's account of spacetime would indeed provide the Argument for Temporal Asymmetry in Spacetime with the support it requires and pose a difficult

objection to my geometrical/material distinction, the view of spacetime suggested by Janssen, which is incompatible with Brown's account, provides us with more satisfying explanations for why the laws behave as they do while still providing explanations for the phenomena Brown charges it can't explain. Thus, we should reject Brown's account, and once again we find that the Argument for Temporal Asymmetry in Spacetime fails.

The failure of the Argument for Temporal Asymmetry in Spacetime may provide us with an important moral. Most obviously, the fact that the fundamental laws of physics are temporally irreversible may not give us any reason to believe that spacetime is temporally anisotropic, and even if it does, the relationship between dynamical laws and spacetime structure may be significantly more complicated than authors like Maudlin and Horwich seem to acknowledge. If the connection between dynamical asymmetries and spacetime structures is to be defended, its defenders must explain what virtues of theory-interpretation pairs that follow the geometrical interpretation of ϕ have that theory-interpretation pairs that follow the material interpretation of ϕ lack and why we should believe that theory-interpretation pairs that exemplify these particular virtues are the ones that most aptly describe the world. I believe my analysis of the Argument for Temporal Asymmetry in Spacetime suggests this will be an uphill battle, but that does not mean it will be an ultimately futile one.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abe, K., et al. (2001), Observation of large CP violation in the neutral B meson system, *Physical Review Letters*, *87*, 091802.

Albert, D. (2000), *Time and Chance*, Harvard University Press.

Arntzenius, F., and H. Greaves (2009), Time reversal in classical electromagnetism, *The British Journal for the Philosophy of Science*, *60*, 557–584.

Baker, D. (2010), Symmetry and the metaphysics of physics, *Philosophy Compass*, *5*, 1157–1166.

Balashov, Y., and M. Janssen (2003), Presentism and relativity, *The British journal for the philosophy of science*, *54*, 327–346.

Belot, G. (2003), Notes on symmetries, in *Symmetries in Physics: Philosophical Reflections*, edited by K. Brading and E. Castellani, pp. 393–411, Cambridge University Press.

Belot, G. (2011), *Geometric Possibility*, Oxford University Press.

Belot, G. (2013), Symmetry and equivalence, in *The Oxford Handbook of Philosophy of Physics*, edited by R. Batterman, pp. 318–339, Oxford University Press.

Brading, K., and E. Castellani (2007), Symmetry in classical physics, in *Handbook of the Philosophy of Physics*, edited by J. Butterfield and J. Earman, pp. 1331–1367, North-Holland.

Brown, H. (2005), *Physical Relativity: Space-Time Structure From A Dynamical Perspective*, Oxford University Press.

Brown, H., and O. Pooley (2006), Minkowski space-time: a glorious non-entity, in *The Ontology of Spacetime*, edited by D. Dieks, pp. 67–89, Elsevier.

Callender, C. (2000), Is time 'handed' in a quantum world?, *Proceedings of the Aristotelian Society*, *100*, 247–269.

Callender, C. (2011), What makes time special, fQXI essay contest submission.

Carroll, S. (2010), *From Eternity to Here: The Quest for the Ultimate Theory of Time*, Penguin Press.

Díaz, J. (2003), Short guide to YaLie. Yet another Lie Mathematica package for Lie symmetries.

Earman, J. (1974), An attempt to add a little direction to "the problem of the direction of time", *Philosophy of Science*, *41*, 15–47.

Earman, J. (1986), *A primer on determinism*, Springer.

Earman, J. (1989), *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*, MIT Press.

Friedman, M. (1983), *Foundations of Space-Time Theories*, Princeton University Press.

Frisch, M. (2011), Principle or constructive relativity, *Studies in History and Philosophy of Modern Physics*, *42*, 176–183.

Greaves, H. (2010), Towards a geometrical understanding of the CPT theorem, *The British Journal for the Philosophy of Science*, *61*, 27–50.

Hamad, M., I. Hassanien, and E. El-Nahary (2011), Discrete symmetries analysis of Burgers equation with time dependent flux at the origin, *World Applied Sciences Journal*, *12*, 2291–2300.

Healey, R. (2009), Perfect symmetries, *The British Journal for the Philosophy of Science*, *60*, 697–720.

Horwich, P. (1987), *Asymmetries in Time: Problems in the Philosophy of Science*, MIT Press.

Hydon, P. (1998a), Discrete point symmetries of ordinary differential equations, *Proceedings of the Royal Society of London A*, *454*, 1961–1972.

Hydon, P. (1998b), How to find discrete contact symmetries, *Journal of Nonlinear Mathematical Physics*, *5*, 405–416.

Hydon, P. (1999), How to use Lie symmetries to find discrete symmetries, *Modern Group Analysis VII, ed. NH Ibragimov, KR Naqvi and E. Straume, MARS Publishers, Trondheim*, pp. 141–147.

Hydon, P. (2000a), *Symmetry Methods for Differential Equations: A Beginner's Guide*, Cambridge University Press.

Hydon, P. (2000b), How to construct the discrete symmetries of partial differential equations, *European Journal of Applied Mathematics*, *11*, 515–527.

Hydon, P. (2001), Discrete symmetries of differential equations, in *The geometrical study of differential equations: NSF-CBMS Conference on the Geometrical Study of Differential Equations, June 20-25, 2000, Howard University, Washington*, vol. 285, pp. 61–70, American Mathematical Society.

Ismael, J., and B. van Fraassen (2003), Symmetry as a guide to superfluous theoretical structure, in *Symmetries in Physics: Philosophical Reflections*, edited by K. Brading and E. Castellani, pp. 371–392, Cambridge University Press.

Janssen, M. (2009), Drawing the line between kinematics and dynamics in special relativity, *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, *40*, 26–52.

Jansson, I. (2011), Explanation and dependence, Ph.D. thesis, The University of Michigan.

Laine-Pearson, F., and P. Hydon (2003), Classification of discrete symmetries of ordinary differential equations, *Studies in Applied Mathematics*, *111*, 269–299.

Lange, M. (1995), Are there natural laws concerning particular biological species?, *The Journal of philosophy*, *92*.

Lange, M. (2007), Laws and metalaws of nature, *Harvard Review of Philosophy*, *15*, 21–36.

Lange, M. (2009a), A tale of two vectors, *Dialectica*, *63*, 397–431.

Lange, M. (2009b), *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*, Oxford University Press.

Levi, D., and M. Rodríguez (2004), Lie discrete symmetries of lattice equations, *Journal of Physics A: Mathematical and General*, *37*, 1711–1725.

Lewis, D. (1983), New work for a theory of universals, *Australasian Journal of Philosophy*, *61*, 343–377.

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

Maudlin, T. (2007), *The Metaphysics Within Physics*, Oxford University Press.

North, J. (2008), Two views on time reversal, *Philosophy of Science*, *75*, 201–223.

North, J. (2009), The "structure" of physics: A case study, *Journal of Philosophy*, *106*, 57–88.

Norton, J. (2008), Why constructive relativity fails, *The British Journal for the Philosophy of Science*, *59*, 821–834.

Olver, P. (1993), *Applications of Lie groups to Differential Equations, 2nd ed*, Springer Verlag.

Pooley, O. (2003), Handedness, parity violation, and the reality of space, in *Symmetry in Physics: Philosophical Reflections*, edited by K. Brading and E. Castellani, pp. 250–280, Cambridge University Press.

Price, H. (1997), *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*, Oxford University Press.

Price, H. (2011), The flow of time, in *The Oxford Handbook of Time*, edited by C. Callender, pp. 276–311, Oxford University Press.

Roberts, B. (2010a), Time reversal in quantum mechanics, preprint.

Roberts, B. (2010b), A foundational approach to time reversal, preprint.

Roberts, J. (2008), A puzzle about laws, symmetries, and measurability, *The British Journal for the Philosophy of Science*, *59*, 143–168.

Rynasiewicz, R. (1986), The universality of laws in space and time, in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 66–75.

Sachs, R. (1987), *The Physics of Time Reversal*, University of Chicago Press.

Silberberg, G. (2005), Discrete symmetries of the black-scholes equation, preprint.

Sklar, L. (1975), Methodological conservatism, *The Philosophical Review*, *84*, 374–400.

Sklar, L. (1977), *Space, Time, and Spacetime*, University of California Press.

Skow, B. (2007), What makes time different from space?, *Noûs*, *41*, 227–252.

Streater, R., and A. Wightman (1980), *PCT, Spin, Statistics, and All That*, Princeton University Press.

Tooley, M. (1977), The nature of laws, *Canadian Journal of Philosophy*, *7*, 667–698.

van Fraassen, B. (1989), *Laws and Symmetry*, Clarendon Press.

Zeh, H. D. (2001), *Physical Basis of the Direction of Time*, Springer.