On Certain *A Priori* Claims in the Foundations of Spacetime Theories

James Read

Hertford College, University of Oxford, OX1 3BW, UK

james.read@philosophy.ox.ac.uk

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

August 2018

कसोरी अडेको छ यो वस्तुसत्ता ?
जगत् सृष्टिमा कालको के महत्ता ?
यिनै प्रश्नलाई समाती समाती
म उक्लन्छु उद्देश्य कैलासमाथि

–मुकुन्द शर्मा

Abstract

I critique three *a priori* claims present in the literature on the foundations of spacetime theories:—

1. Spacetime structure has is chronogeometric significance necessarily.
2. The only candidates for representing physical quantities are objects which can be specified in a coordinate-independent manner.
3. Symmetry- and duality-related models of physical theories may be regarded *ab initio* as being physically equivalent.

Acknowledgements

I would like to thank the following individuals and organisations:

1. My parents, Clare and Jonathan Read; grandparents, Caryl and Michael Baron and Meg and John Read; wife, Sumana Chalise; and parents-in-law, Mayadevi and Mukunda Sharma Chalise, for their love and support.

2. My supervisors, Harvey Brown and Oliver Pooley, for their superlative guidance.

3. The Arts and Humanities Research Council; Hertford College, Oxford; and the University of Illinois at Chicago, for generous financial support.

4. Jeremy Butterfield, Adam Caulton, Erik Curiel, Sebastian De Haro, Neil Dewar, Patrick Dürr, Ricardo Heras, Nick Huggett, Eleanor Knox, Baptiste Le Bihan, Dennis Lehmkuhl, Niels Linnemann, Niels Martens, Keizo Matsubara, Tushar Menon, Thomas Møller-Nielsen, Wayne Myrvold, Simon Saunders, Teruji Thomas, James Owen Weatherall, and Christian Wüthrich, for feedback on parts or all of this thesis.

5. Augustin Baas, Dave Baker, Gordon Belot, Richard Dawid, Sean Gryb, Josh Hunt, Vincent Lam, Martin Lesourd, Brian Pitts, Dean Rickles, Laura Ruetsche, Nic Teh, Karim Thébault, and David Wallace, for valuable discussions.

Contents

Introduction

In this thesis, I critique three *a priori* claims widespread in the literature on the foundations of spacetime theories.

Part I

Claim: *The geometrical object representing spacetime structure in the theory under consideration has its chronogeometric significance of necessity.*

Part I consists in three chapters. In Chapter 1, I first present cases in which 'minimal coupling' (i.e., a restriction the allowed class of matter fields in general relativity (GR)) is incompatible with the 'strong equivalence principle' (i.e., the condition, roughly speaking, that dynamical equations governing non-gravitational fields in GR should be locally 'special relativistic')—*pace* widespread assumptions to the contrary. I precisify the strong equivalence principle (SEP) in four ways, before drawing upon this work in order to identify two unexplained coincidences—or 'miracles'—in the foundations of GR. Having done so, I argue that while a 'geometrical' view, according to which the metric field of GR has its chronogeometric significance of necessity, would be able to account for these 'miracles', there are good reasons to question such a view.

Chapter 2 constitutes the 'positive' sequel to the above: I demonstrate that the two 'miracles' of GR admit of natural explanation in one candidate successor theory to GR—*viz*., perturbative string theory. There exists a parallel here with other explanations of coincidences in historical spacetime theories on moving to successor theories—e.g., the explanation of the coincidence of gravitational and inertial masses in Newtonian gravitation theory (NGT) on moving to GR.

In Chapter 3, I consider the debate between the 'dynamical' versus 'geometrical' approaches to spacetime theories in more detail, and argue that, while the claim that the metric field has its chronogeometric significance of necessity is indeed false, there exists an alternative formulation of the geometrical approach which does *not* make this claim. Thus, the geometrical approach should not be written off prematurely by advocates of the dynamical approach.

Part II

Claim: *The only candidates for representing physical quantities are objects which can be specified in a coordinate-independent manner.*

In Part II—which consists of one chapter, Chapter 4—I address the (widespread) claim that the only candidates for representing physical quantities are objects which can be specified in a coordinate-independent manner. I do so obliquely, via reconsideration of an ongoing debate regarding the status of gravitational energy in GR. By deploying functionalist criteria for the definition of physical quantities, I argue that not only does there exist a frame-independent notion of gravitational energy in some models of GR, but also that one may define a frame-dependent notion of gravitational energy, in *all* models of GR. I am not alone in regarding frame-dependent notions as physical; for example, the moral of Chapter 4 is consonant with Rovelli's writing that

"Gauge quantities cannot be predicted, but can often be measured. They measure our localization in time, our orientation in general relativistic space, in the internal space, and similar." [189, p. 101]

Part III

Claim: *Symmetry- and duality-related models of physical theories may be regarded* ab initio *as being physically equivalent.*

In Part III, I address the question of whether symmetry- and duality-related models of physical theories may be regarded *ab initio* as being physically equivalent. Though there exists a large body of literature endorsing such a claim, I resist it, and argue for an alternative: symmetry- and duality-related models may *only* be regarded as being physically equivalent once we have to hand a perspicuous account of the common physical reality underpinning that equivalence.

The Part divides into two chapters; both are concerned largely with such issues in the interpretation of *dualities*, though they draw upon parallel debates regarding the interpretation of symmetry transformations. In Chapter 5, I address the question of in what that common physical reality is supposed to consist in the case of dual theories; the chapter should also serve as an introduction to issues in the interpretation of duality-related theories more generally. In Chapter 6, I argue against the claim that symmetry- and duality-related models of physical theories may be regarded *ab initio* as being physically equivalent.

Publication information

All chapters in this thesis are based upon papers which are either forthcoming in peer-reviewed collections or journals, or which are under review. The publication information for the paper associated with each chapter is as follows:

Chapter 1: James Read, Harvey R. Brown and Dennis Lehmkuhl, "Two Miracles of General Relativity", Studies in History and Philosophy of Modern Physics, 2018. (Forthcoming.)

Chapter 2: James Read, "On Miracles and Spacetime", 2018. (Under review.)

Chapter 3: James Read, "Explanation, Geometry, and Conspiracy in Relativity Theory", in C. Beisbart, T. Sauer and C. Wüthrich (eds.), *Thinking About Space and Time: 100 Years of Applying and Interpreting General Relativity*, vol. 15 of the *Einstein Studies* series, Basel: Birkhäuser, 2018. (Forthcoming.)

Chapter 4: James Read, "Functional Gravitational Energy", British Journal for the Philosophy of Science, 2018. (Forthcoming.)

Chapter 5: James Read and Baptiste Le Bihan, "Duality and Ontology", Philosophy Compass, 2018. (Forthcoming.)

Chapter 6: James Read and Thomas Møller-Nielsen, "Motivating Dualities", Synthese, 2018. (Forthcoming.)

Permission to include material based upon jointly-authored work has been obtained from the Faculty of Philosophy, University of Oxford, as well as from all relevant co-authors. I am the lead author on all of the above works, and the author of the entirety of this thesis. Due acknowledgement has been given over the course of this thesis to the contributions of my collaborators.

This thesis is approximately 61,000 words long.

Part I

Dynamical and Geometrical Approaches to Spacetime

Chapter 1

Two Miracles of General Relativity

I approach the physics of *minimal coupling* in general relativity, demonstrating that in certain circumstances this leads to (apparent) violations of the *strong equivalence principle*, which states (roughly) that, in general relativity, the dynamical laws of special relativity can be recovered at a point. I then assess the consequences of this result for the *dynamical perspective on relativity*, finding that potential difficulties presented by such apparent violations of the strong equivalence principle can be overcome. Next, I draw upon my discussion of the dynamical perspective in order to make explicit two 'miracles' in the foundations of relativity theory. I close by arguing that the above results afford insight into the nature of special relativity, and its relation to general relativity.

1.1 Introduction

Recently, there has arisen a heightening of interest in the physics community in the coupling of Maxwell electrodynamics to Einstein gravity. For example, the *mini-*

mal coupling of electromagnetism to gravity leads to curvature terms in second (and higher) order dynamical equations governing non-gravitational fields, written at any point;[1] however, there also exist different possible coupling schemes, according to which one recovers *different* dynamical equations at any point. In literature such as [93, 105], authors have attempted to elaborate the physical and mathematical details of these different possible coupling schemes.

The purpose of this chapter is to pursue a *philosophical* branch of enquiry into the coupling of electromagnetism to gravity; this field provides fertile ground for foundational insights into both special and general relativity (SR and GR, respectively). A crucial starting point in this regard is that the existence of curvature terms in higher order dynamical equations governing non-gravitational fields, written at any point, indicates violations of the *strong equivalence principle* (sometimes: *Einstein equivalence principle*[2]). In this thesis, I take the basic idea of the principle to be that (what are typically understood to be) the dynamical laws of SR—i.e. dynamical laws governing matter fields in a fixed Minkowski background, featuring no curvature terms—are recovered in GR at any point. This observation that there exist apparent violations

[1] In this thesis, I mean by 'matter fields', or 'non-gravitational fields', those for which there exists an associated stress-energy tensor, and by 'gravitational fields' those for which there exists no such stress-energy tensor—this distinction is in the spirit of [121]. In the context of general relativity, this means that the metric field is identified as a gravitational field (see [42] for a proof against the existence of a tensorial expression of such stress-energy), whereas all other fields typically of interest (e.g. Klein-Gordon fields, electromagnetic fields, etc.) are matter fields. There exist subtle issues regarding 'gravitational' stress-energy in general relativity—see e.g. [95, 118, 178] and Ch. 4 for further discussion. Note also that this distinction between matter and gravitational fields may break down in the case of other spacetime theories—for example, in Newtonian gravitation theory, it is possible to define a stress-energy tensor associated with the potential φ, in spite of this field naturally being regarded as 'gravitational' (cf. [58]). Nevertheless, for my purposes, the above distinction will suffice. (I am grateful to Dennis Lehmkuhl for suggesting that the distinction between matter and gravitational fields be articulated in this manner.)

[2] See e.g. [29, §IV], and [164, p. 219]. I distinguish different ways of making this principle precise in §1.3.2. Despite the fact that the principle has been named after Einstein, it should be noted that Einstein himself meant something different when he spoke of the 'equivalence principle'. Though he subscribed both to his definition of the equivalence principle, and to the local validity of special relativity, he saw the two principles as clearly distinct, as is most clear in [72]. For details of what Einstein calls the 'equivalence principle', see [152]. For these reasons, I use the 'strong equivalence principle' nomenclature in this thesis. (For more on different versions of the equivalence principle in general, see [123].)

of the strong equivalence principle (made by Eddington as early as 1923 [66, p. 176]) requires a rethinking of the foundations of GR, as this principle is still sometimes asserted to be universally valid in texts on the subject (see e.g. [24, p. 169]).

To illustrate, the details of minimal coupling raise important questions regarding the *dynamical perspective on relativity* (principally advanced in [24, 27, 28]), according to which the metric field in GR, though ontologically distinct from matter fields, acquires its *chronogeometric significance* via the way it couples to matter fields in accord with the strong equivalence principle, and codifies the symmetries of local dynamical laws. The claim that there exist violations of the strong equivalence principle raises questions over whether this account can still go through. Though the subtleties presented by violations of this principle require one to make slight modifications to the dynamical view in the context of GR, I argue that the problems for the dynamical perspective presented by such cases can be overcome.

Having appreciated the details of such cases, other foundational results follow. In particular, consideration of how the dynamical perspective should be understood in GR in light of these results leads me to identify two crucial foundational aspects—or 'miracles'—in the context of GR.[3] In turn, this leads to an understanding of the relation of SR to GR which extends the previous notion that dynamical laws in a Minkowski background being recovered in GR at any point characterises this relation—instead, I suggest a liberalised notion of what constitutes a 'special relativistic' theory, according to which the sole defining characteristic of such theories is the *Poincaré invariance* of dynamical laws governing matter fields.[4]

[3] The articulation of these two 'miracles' came out of fertile discussions with Harvey Brown in 2015—the nomenclature is his. By 'miracle', I mean something surprising or puzzling, the explanation of which remains outstanding. Another example of such a 'miracle' in physics would be the proportionality of gravitational and inertial masses in Newtonian mechanics. Just as, in that case, one has to move to a successor theory (namely, GR) to account for such a 'miracle' (cf. [219]), in my view accounting for the two 'miracles of GR' which I specify in this chapter will require recourse to considerations from physical theories external to GR—see Ch. 2.

[4] The distinction between 'dynamical laws in a Minkowski background' and 'Poincaré-invariant

The structure of this chapter is as follows. In §1.2, I discuss electromagnetism both in fixed Minkowski spacetime, and in dynamical, generically curved spacetime—the latter theory constructed on the basis of minimal coupling. In §1.3, I consider the ramifications of the minimal coupling of electromagnetism to gravity for our understanding of the strong equivalence principle. In §1.4, I expound potential problems for the dynamical perspective which arise from this work, showing that ultimately such worries can be overcome. In §1.5, I highlight two 'miracles' in the foundations of GR. In §1.6, I raise concerns about the 'geometrical' alternative to the dynamical approach in the context of GR—in particular, about those versions of the geometrical approach committed to the claim that the metric field has its chronogeometric significance of necessity (cf. Ch. 3 for alternative versions of the geometrical approach, which do not commit to this claim). In §1.7, I make some remarks on the nature of SR and its relation to GR. I close the chapter in §1.8 with conclusions and outlook.

1.2 Electromagnetism

1.2.1 Minkowski Spacetime

Consider Maxwell electrodynamics in a flat, static Minkowski background. In that situation, the metric field η_{ab} is a *fixed field* in the sense of [171, p. 13] (i.e. η_{ab} is identically the same in all kinematically possible models), and the Maxwell equations can be written:[5]

dynamical laws' will be made clear in the body of this chapter.

[5]Throughout this thesis, abstract (i.e. coordinate-independent) indices are written in Latin script beginning $a, b, c, \ldots$; indices in a coordinate basis are written in Greek script; 3-vector indices in a coordinate basis are written in Latin script beginning $i, j, k, \ldots$; semicolons indicate covariant derivatives; commas indicate partial derivatives; and the Einstein summation convention is used. Round brackets around indices denote symmetrisation over those indices; square brackets around indices denote antisymmetrisation. In addition, I set $\epsilon_0 = \mu_0 = c = G_N = 1$.

$$F^{ab}{}_{;b} = J^a, \quad (1.2.1)$$

$$F_{[ab;c]} = 0, \quad (1.2.2)$$

where covariant derivatives are here taken with respect to the torsion-free derivative operator compatible with η_{ab} (see e.g. [128, p. 49]); $F_{ab} = -F_{ba}$ is the Faraday tensor; and I define

$$F^{ab} := \eta^{ac}\eta^{bd}F_{cd}, \quad (1.2.3)$$

with η^{ab} the inverse of η_{ab}. (1.2.1) can be derived from a variational principle: the standard Lagrangian density for a sourced Maxwell field is given by

$$\mathcal{L}_{\text{EM}} = -\frac{1}{4}F_{ab}F^{ab} - A_a J^a, \quad (1.2.4)$$

where A^a is the electromagnetic vector potential; $A_a := \eta_{ab}A^b$; and J^a is a source term. Recalling that F_{ab} can be written as

$$F_{ab} = 2A_{[a;b]}, \quad (1.2.5)$$

and then applying Hamilton's principle with respect to arbitrary variations in the A_a, one obtains (1.2.1). (1.2.2) follows from (1.2.5).[6] The stress-energy tensor for the

[6] If the derivative operator in (1.2.1) and (1.2.2) is not assumed to be metric-compatible, then a metric field is only required to write down (1.2.1), and not (1.2.2). Cf. appendix 1.A.

electromagnetic field is[7]

$$T^{ab} = F^{ac}F^{b}_{\ c} - \frac{1}{4}\eta^{ab}F_{cd}F^{cd}. \tag{1.2.6}$$

Finally, F_{ab} satisfies a *wave equation*, which can be derived by manipulation of (1.2.1) and (1.2.2):

$$F_{ab;c}{}^{c} = 2J_{[a;b]}. \tag{1.2.7}$$

1.2.2 Dynamical Spacetime

In GR, the metric field g_{ab} is not fixed in all kinematically possible models, and moreover is dynamically coupled to the matter fields via Einstein's equation. Additionally, one need not restrict *ab initio* the allowed form of the dynamical equations for matter fields (e.g., arbitrary contractions of the matter fields and the curvature tensor $R^{a}{}_{bcd}$ associated with the torsion free, metric-compatible derivative operator ∇_a are permitted). However, that being said, one often restricts to dynamical equations for non-gravitational fields which are *minimally coupled*. The prescription for constructing such minimally coupled equations is the following:

> Minimally coupled dynamical equations for matter fields in GR are constructed from dynamical equations for matter fields featuring coupling to a fixed Minkowski metric field η_{ab} and no curvature terms, by replacing all instances of η_{ab} with a generic Lorentzian metric field g_{ab}, and replacing all

[7]Recall that the stress-energy tensor is defined through $T^{ab} = \frac{2}{\sqrt{g}}\frac{\delta S}{\delta g_{ab}}$, where g is the metric determinant, and S is the action to which the matter Lagrangian—here $\mathcal{L}_{\text{EM}}$—is associated.

instances of the torsion-free derivative operator compatible with η_{ab} with the torsion-free derivative operator compatible with g_{ab}.

In the case of electromagnetism in the GR context, if one couples the F_{ab} to g_{ab} via

$$F^{ab}{}_{;b} = J^a, \tag{1.2.8}$$

$$F_{[ab;c]} = 0, \tag{1.2.9}$$

and

$$G^{ab} := R^{ab} - \frac{1}{2}g^{ab}R = 8\pi T^{ab}, \tag{1.2.10}$$

then one obtains Maxwell electrodynamics minimally coupled to Einstein gravity. Here, T^{ab} is the stress-energy tensor of F_{ab}; g^{ab} is the inverse of g_{ab}; indices are lowered with respect to g_{ab} and raised with respect to g^{ab}; the derivative operator is torsion-free and metric-compatible; and the Ricci tensor $R_{ab} := R^c{}_{acb}$ and Ricci scalar $R := g^{ab}R_{ab}$ are those associated to that derivative operator. (1.2.8) and (1.2.9) are Maxwell's equations in this (dynamical and generically curved) spacetime; and (1.2.10) is Einstein's equation. In this case, T^{ab} on the right hand side of (1.2.10) reads[8]

$$T^{ab} = F^{ac}F^b{}_c - \frac{1}{4}g^{ab}F_{cd}F^{cd}. \tag{1.2.11}$$

In this minimally-coupled Maxwell-Einstein system, the wave equation for F_{ab}

[8] More generally, T^{ab} is the stress-energy tensor associated to *all* matter fields in the theory.

becomes, after using (1.2.8), (1.2.9), and Ricci-like identities,[9]

$$F_{ab;c}{}^{c} = 2\left(F^{e}{}_{[b}R_{a]e} - R_{abcd}F^{cd} + J_{[a;b]}\right). \tag{1.2.12}$$

In (1.2.7), the two curvature terms on the right hand side of (1.2.12) were not present, as the metric field was fixed to be η_{ab}—*a fortiori* flat—in all kinematically possible models. Dropping this assumption leads to the generalised form (1.2.12).

There is an important feature of second-order equations such as (1.2.12), valid in minimally-coupled Maxwell-Einstein dynamics. To see this, first recall that in a coordinate basis $\{e_\mu\}$, the *connection components* $\Gamma^{\mu}{}_{\nu\rho}$ associated to a derivative operator ∇_a are defined by $\nabla_\rho e_\nu =: \Gamma^{\mu}{}_{\nu\rho} e_\mu$. Then, where M is the spacetime manifold, at any $p \in M$ we can choose *normal coordinates*, such that $\Gamma^{\mu}{}_{(\nu\rho)}(p) = 0$ in those coordinates; for a torsion-free derivative operator, we can in fact choose normal coordinates such that $\Gamma^{\mu}{}_{\nu\rho}(p) = 0$. (Note that the connection components away from p will in general *not* vanish.) If the unique torsion-free, metric compatible derivative operator is used, then in normal coordinates we also have $g_{\mu\nu,\rho}(p) = 0$, and we may find a subclass of normal coordinates at p such that $g_{\mu\nu}(p) = \mathrm{diag}(-1,1,1,1)$.[10] Since $g_{\mu\nu}(p)$ takes this diagonal form—preserved under Poincaré transformations—one might write $g_{\mu\nu}(p) = \eta_{\mu\nu}$ (cf. e.g. [141, p. 1055]). This notwithstanding, however, any claim to the effect that the metric field 'reduces' to the Minkowski metric at p in normal coordinates should be met with suspicion—for in general, second (and higher) order derivatives of the metric field *do not* vanish at p, in these coordinates.

[9] I.e. identities of the form $(\nabla_c\nabla_d - \nabla_d\nabla_c)T^{a_1\ldots a_k}{}_{b_1\ldots b_l} = R^{a_1}{}_{ecd}T^{e\ldots a_k}{}_{b_1\ldots b_l} + \cdots + R^{a_k}{}_{ecd}T^{a_1\ldots e}{}_{b_1\ldots b_l} - R^{e}{}_{b_1cd}T^{a_1\ldots a_k}{}_{e\ldots b_l} - \cdots - R^{e}{}_{b_lcd}T^{a_1\ldots a_k}{}_{b_1\ldots e}$.

[10] I say here 'a subclass', for the conditions on normal coordiantes at p are preserved under all affine transformations, whereas the condition $g_{\mu\nu}(p) = \mathrm{diag}(-1,1,1,1)$ is preserved only under a subclass of such transformations—*viz.*, Poincaré transformations. In the remainder of this chapter, I invariably focus upon those normal coordinates in which this diagonalisation condition holds, without explicit qualification.

This point will be of importance in what follows.

In normal coordinates at p, connection components vanish—and elements of the class of frames in which such is the case, and in which $g_{\mu\nu}(p) = \mathrm{diag}(-1, 1, 1, 1)$, are related to one another by Poincaré transformations. Moreover, first order, minimally coupled dynamical equations such as (1.2.8) and (1.2.9) do not feature curvature terms at p. In this way, first order, minimally coupled dynamical equations recover their 'Minkowski background' form at p. However, this is *only* generically true of first order equations; equations that are second or higher order may contain curvature pieces and higher order covariant derivatives—consider, for example, (1.2.12) versus (1.2.7). Since curvature is represented by the Riemann *tensor*, this object cannot be made to vanish at p. This implies that in GR it is not always the case that laws recover their original 'Minkowski background' forms at a point. Another way to see this is to note that by expanding out the curvature and covariant derivative terms in such equations, we obtain terms containing derivatives of connection coefficients, which in general cannot be made to vanish at p in normal coordinates.[11]

1.3 Minimal Coupling and the Equivalence Principle

1.3.1 Local Dynamical Equations

What should one make of the observation that minimal coupling yields second order equations such as (1.2.12) containing curvature terms which do not vanish at a point? To answer this question, it is useful to distinguish two forms that the dynam-

[11] Cf. [29, §IV]. The present discussion goes further than that paper in several respects—for example, in explicitly delineating precisifications of the strong equivalence principle which are in tension with the above results.

ical laws for non-gravitational fields might take in normal coordinates at any given point $p \in M$ in GR: (a) Poincaré invariant, with no terms containing the Riemann tensor or its contractions; or (b) Poincaré invariant simpliciter. In §1.2, we saw that minimal coupling yields violations of the claim that, at any $p \in M$, dynamical laws take a Poincaré invariant form with no terms featuring the Riemann tensor or its contractions (i.e., form (a)). However, this does not necessarily imply violations of the analogous claim regarding such laws taking, at any $p \in M$, a Poincaré invariant form simpliciter (i.e., form (b)).

One important point to make regarding minimal coupling is the following: although a certain form of dynamical law (i.e., those laws featuring coupling to a fixed Minkowski metric field, and containing no curvature terms) is used to determine the class of minimally coupled dynamical laws for non-gravitational fields in GR, there exists no *a priori* restriction on the form that these general relativistic, minimally coupled equations take locally, at any $p \in M$—in particular, there exists no restriction that we recover locally the dynamical laws from which we began. That said, it is straightforward to show that such minimally coupled dynamical equations in GR take a Poincaré invariant form—but *not* necessarily a form in which terms containing the Riemann tensor or its contractions vanish—at any $p \in M$.[12] In this way, one can consider (minimal) coupling to matter to *restrict* the local symmetry group of the dynamical equations of GR to the Poincaré group.

Some further words on the ambiguity of the application of minimal coupling are in order. Clearly, the minimal coupling prescription presented in §1.2.2 does not specify the dynamical equations to which the procedure is to be applied. Though in §1.2.2,

[12] Cf. appendix 1.A. The work presented in that appendix also brings out when such local dynamical equations are invariant under a *broader* symmetry group than the Poincaré group. The reader's attention should also be drawn to §1.6 and appendix 1.B, in which cases are presented in which dynamical equations for matter fields take a yet simpler form (in the sense that further terms in these equations vanish) in a class of frames related by a *subgroup* of the Poincaré group.

the procedure was applied to (1.2.1) and (1.2.2), yielding (respectively) (1.2.8) and (1.2.9), which in turn were used to construct (1.2.12), it is *not* the case that, were the minimal coupling procedure applied directly to (1.2.7), (1.2.12) would have been obtained. Thus, there is a sense in which the minimal coupling prescription is ambiguous.[13] This is in line with Goenner's observation that "such a procedure is unique only for *first*-order partial differential equations" [87, p. 866].

A related point here is the following. One might think that it is no surprise that equations such as (1.2.12) do not reduce to (1.2.7) (i.e., to a Poincaré invariant form, with no curvature terms) in normal coordinates at any $p \in M$, for (1.2.12) was not obtained *directly* via minimal coupling. This thought, however, is ultimately by-the-by, for even if minimal coupling were applied directly to (1.2.7), to obtain

$$F_{ab;c}{}^{c} = 2J_{[a;b]}, \tag{1.3.1}$$

where now index contraction is taken with respect to a generic Lorentzian g_{ab} satisfying (1.2.10), and the unique torsion free derivative operator ∇_a compatible with g_{ab} is used, (1.3.1) would *still* not take at any $p \in M$ and in normal coordinates a Poincaré invariant form with no curvature terms—the reason being that, in a coordinate basis, (1.3.1) features derivatives of connection components, which cannot be made to vanish in normal coordinates.

1.3.2 The Strong Equivalence Principle

The work presented in §1.2 has important ramifications for the *strong equivalence principle*. I here introduce two distinct formulations of this principle, which I call **EP1** and

[13] For discussion related to this point, see [141, §16.3].

EP2:[14]

EP1: The dynamical equations for non-gravitational fields reduce to a Poincaré invariant form, with no terms featuring the Riemann tensor or its contractions, at any $p \in M$.

EP2: The dynamical equations for non-gravitational fields reduce to a Poincaré invariant form at any $p \in M$.

The results of §1.2 indicate that **EP1** is incompatible with minimal coupling for higher-order dynamical equations like (1.2.12). Claims to the contrary can, however, be found in the foundational literature; consider for example the following quote, given by Brown at [24, pp. 170-171]:

> [Minimal coupling involves the] claim that the matter fields do not couple to the Riemann curvature tensor or its contractions. Recall that in SR, inertial frames are global, which implies that the curvature vanishes everywhere, and hence trivially makes no appearance in the laws of physical interactions. This feature is now absorbed into GR in the requisite local context.

Since terms featuring in (1.2.12) contain the Riemann *tensor*, these cannot be made to vanish at a point; this is in manifest contradiction with the above quote. In fact, second order equations such as (1.2.12) constitute a straightforward counterexample to any claims to the effect that **EP1** may universally be regarded as holding when dynamical equations for matter fields are constructed via minimal coupling. Note,

[14] Sometimes, **EP1** and **EP2** are referred to as 'pointy' versions of the strong equivalence principle—for discussion, see [85, 111, 155].

however, that *insofar as curvature terms may be ignored at $p \in M$*, **EP1** may be regarded as holding at p. Whether this is the case will principally depend upon the strength of curvature effects relative to the experimental apparatuses available; if one's experimental apparatuses are insensitive to such effects, then the principle may be regarded as holding approximately at p for the matter field measured by those apparatuses—see below for further discussion. By contrast to **EP1**, note that curvature couplings in higher-order minimally coupled dynamical equations for matter fields do not pose problems for the validity of **EP2** in the context of matter fields obeying such dynamical equations—indeed, the results of §1.A demonstrate that there is no tension between **EP2** and minimal coupling.

Recall now Brown's discussion of the strong equivalence principle at [24, p. 169]:[15]

> There exists in a neighbourhood of each event preferred coordinates, called *locally inertial* at that event. For each fundamental non-gravitational interaction, to the extent that tidal gravitational forces can be ignored, the laws governing the interaction find their simplest form in these coordinates. This is their *special relativistic form*, independent of spacetime location.

To understand this quote, it is useful to extend **EP1** and **EP2** to the *neighbourhood* of any $p \in M$:

EP1': The dynamical equations for non-gravitational fields reduce to a Poincaré invariant form, with no terms featuring the Riemann tensor or its contractions, in a neighbourhood of any $p \in M$.

[15] Other similar presentations of the strong equivalence principle can be found in e.g. [111, §3.4] and [112, p. 874].

EP2': The dynamical equations for non-gravitational fields reduce to a Poincaré invariant form in a neighbourhood of any $p \in M$.

As with **EP1**, whether **EP1'** holds approximately will principally depend upon the strength of curvature effects relative to the experimental apparatus available. (Arguably, however, such assumptions are more plausible in the context of **EP1'** than of **EP1**, since experimental apparatuses always occupy some extended region of spacetime.)[16] In this case, however, the size of the neighbourhood is also relevant: in a larger neighbourhood of p, one is more likely to be able to detect curvature effects with one's experimental apparatus—in which case, **EP1'** will not hold as an approximate principle. To illustrate, consider Pound-Rebka experiments to detect gravitational redshift.[17] The fact that these experimental apparatuses are not sensitive to curvature effects in their neighbourhood means that **EP1'** can be regarded as holding in that neighbourhood for the fields that those apparatuses measure; in light of this, a flat spacetime account of the Pound-Rebka results can be offered. For further details, see [29, §III]. (Though **EP2'** is principally introduced by analogy with **EP2**, it is important to note that, like **EP1'**, this principle also holds only approximately in a neighbourhood for minimally coupled dynamical equations—for it relies upon the metric field g_{ab} to which the matter fields couple being approximately Minkowskian in a neighbourhood.)

Clearly, it is important that one be precise about what is meant by 'tidal gravitational forces' in quotations such as that above. In the context of dynamical equations for non-gravitational fields in the neighbourhood of some $p \in M$, I understand 'terms representing tidal forces' to mean terms in those equations featuring the Riemann tensor or its contractions. As a result, in contexts in which terms representing tidal forces can be ignored, **EP1'** and **EP2'** hold.

[16] In this regard, cf. [157, §1.9], and discussion at [24, p. 170] and [112, p. 875].
[17] For the original Pound-Rebka paper, see [172].

1.4 The Dynamical Perspective

With this understanding of the limitations to the holding of **EP1** in hand, I now consider potential ramifications of this result for positions in the foundations of relativity theory. To that end, I assess in this section the consequences of the above work for the *dynamical perspective on relativity*, as presented in [24, 27, 28]. I argue that difficulties raised by violations of **EP1** do not present damning problems for this view in the context of GR.

1.4.1 Explanation and Codification

To begin the discussion, I must clarify what is meant by the *dynamical perspective on relativity*. (This position will be elaborated in more depth in Ch. 3.) I first focus *specifically on the case of SR*, and distinguish two positions on the nature of the Minkowski metric field:

(A) The Minkowski metric field is an ontologically distinct and primitive entity; its presence can explain certain facts about the dynamical laws governing matter fields (namely, the fact that these laws are Poincaré invariant).

(B) The Minkowski metric field is not an ontologically distinct and primitive entity; rather, it is a codification of certain facts about the dynamical laws governing matter fields (namely, the fact that these laws are Poincaré invariant).

To endorse (B) is to endorse the *dynamical perspective*; the orthodox line is (A)—sometimes referred to as the *geometrical perspective*. In advocacy of the (A)-view, Maudlin writes:[18]

[18] I take Maudlin's speaking of Minkowski spacetime *in vacuo* to indicate his commitment to the view

If we accept that in a vacuum there is no physical structure, except for the structure of space-time itself, then the behaviour of light in a vacuum implies that *the geometry of spacetime alone determines the trajectory of the light rays*. That is, given any point in the space-time p, the structure of space-time ought to fix where light emitted from that p (in any possible direction) will go. [137, p. 68] (Emphasis in original.)

... the Minkowski geometry takes exactly the same form described in [any] Lorentz coordinate system (by the symmetry of Minkowski spacetime), and the laws of physics take exactly the same coordinate-based form when stated in a coordinate-based language in any Lorentz coordinate system (because the laws can only advert to the Minkowski geometry, and it has the same coordinate-based description). [137, pp. 117-118]

By contrast, the (B)-view (principally advocated by Brown and Pooley—see [24, 27, 28]) is expressed in passages such as the following:[19]

The appropriate structure is Minkowski geometry *precisely because* the laws of the non-gravitational interactions are Lorentz covariant. [24, p. 133]

The relative merits of (A) and (B) in the context of SR will be discussed further in

that the metric field of SR is primitive, and ontologically autonomous of matter fields. Other advocates of the (A)-view arguably include e.g. Friedman [81]. A related view is that of Janssen [12, 108], according to whom one makes a 'common origin inference' from the Poincaré invariance of the dynamical laws to Minkowski spacetime—understood as an expression of the *universal* Poincaré invariance of *all* dynamical laws. On this position, the Minkowski metric field is not necessarily an ontologically autonomous entity. For further discussion as to how Janssen's views align with the (A)- and (B)-views, see [1].

[19] Note that what Brown calls 'Lorentz covariance' here is what I call 'Poincaré invariance' above. Related to this, it is worth noting that arguably the essence of SR is more closely tied to Lorentz- than Poincaré-invariance, for both pre-relativistic and relativistic dynamical laws governing matter fields are locally invariant under translations—it is only the linear part of the class of affine transformations under which such laws are invariant that changed on the transition to relativistic physics.

Ch. 3. For the time being, I sketch how the debate between these two camps changes on moving to GR. The first key difference in the case of GR is that the advocate of (B) concedes that the metric field in this context is an autonomous agent, ontologically distinct from the matter fields of the theory.[20] Hence, for the extensions to GR of *both* (A) and (B), the metric field is *not* reducible to properties of the matter fields.

Though advocates of (A) and (B) agree on the ontological autonomy of the metric field in GR, they may disagree on its relation to *chronogeometry*. For the latter, the dynamics of the metric field tell us that it is 'just another field': "Nothing in the form of the equations *per se* indicates that g_{ab} is the metric of space-time, rather than a $(0, 2)$ symmetric tensor which is assumed to be non-singular" [24, p. 160].[21] How, then, does the metric field attain its *chronogeometric significance* in GR? For the proponent of (B), the metric field "earns its spurs by way of the strong equivalence principle" [24, p. 151]. The reasoning here can be stated explicitly as follows:

1. In the neighbourhood of any $p \in M$, assuming that one's experimental apparatuses are such that terms featuring the Riemann tensor or its contractions can be ignored, we recover **EP1'**, for the fields measured by those apparatuses.[22]

2. In addition, in such a scenario the dynamical metric field of GR manifests Poincaré symmetries—for in this case, higher derivatives of the metric field can be regarded as vanishing.[23]

[20]To claim that the metric field is reducible to the matter fields in GR is to endorse a certain form of *relationism* about the metric field; there are profound difficulties with implementing this programme in GR. An obvious illustration of this difficulty can be found in the existence of vacuum solutions in the theory.

[21]Notation in this quotation has been modified for consistency with the present chapter. For discussion of coordinate-dependent versus -independent approaches in the foundations of spacetime theories, see [217].

[22]Note that this invocation of approximation means that, strictly speaking, we are dealing with the *approximate* chronogeometric significance of the metric field in GR.

[23]By imposing the condition that terms representing tidal forces may be ignored, derivatives of the metric field may be ignored, so this field may indeed be treated as the Minkowski metric field in the

3. The symmetries of the dynamical laws governing non-gravitational fields in the appropriate local neighbourhood of p therefore coincide with the symmetries of the dynamical metric field in this neighbourhood.[24]

4. As a result, inertially moving, stable physical rods and clocks survey this primordial metric field, and so afford it its *chronogeometric significance*, in the neighbourhood of p.[25]

In summary, the point may be put as follows:

> It is because of minimal coupling and local Lorentz covariance that rods and clocks, built out of the matter fields which display that symmetry, behave as if they were reading aspects of the metric field and in so doing confer on [the metric] field a geometrical meaning. [24, p. 176]

It is worth making some comments on the (B)-view. First, the advocate of this approach errs insofar as she universally endorses **EP1** in conjunction with minimal coupling, for we have seen cases in which the minimal coupling prescription applied to certain dynamical laws implies the negation of **EP1** (cf. §1.3). This tension

relevant local vicinity of p. Note that thus far we have understood 'symmetries' to mean *dynamical symmetries*—i.e. transformations upon dynamical equations which leave their form invariant. When speaking of the (local) symmetries of the *metric field*, we are concerned with *spacetime symmetries*, which Pooley defines as "groups of transformations that *preserve spatiotemporal structure* (as encoded in coordinate systems)" [169, §3.1]. (Clearly, the advocate of the dynamical approach may resist the *ab initio* identification of the metric field with 'spacetime'; this, however, does not affect the above presentation of the dynamical view.) I will make more of the distinction between these two classes of symmetries in Ch. 3.

[24]On this point, one might ask, "Since this notion of symmetry coincidence is symmetrical, why not say that the metric field surveys rods and clocks?" My response is that the asymmetry is broken through the fact that stable rods and clocks can be built from matter fields (though cf. footnote 25), which then read off intervals as given by the metric field—but it is not (it appears) the case that stable rods and clocks can be built from the metric field, which then survey matter fields in the theory. I concede, however, that this latter possibility is of interest, and deserves further exploration.

[25]Note that the local coincidence of metric and dynamical symmetries, as delivered by the strong equivalence principle, is only a *necessary* condition for the metric field to have chronogeometric significance *tout court*. For example, as flagged above, the existence of stable rods and clocks built from matter fields is an additional, non-trivial assumption.

is not necessarily problematic, however, for what is sufficient for the (B)-view to go through in the general relativistic context (for the reasons detailed above) is that, in an appropriate neighbourhood of any $p \in M$, terms featuring the Riemann tensor or its contractions may be ignored—so that **EP1'** holds in this neighbourhood.[26]

What is a plausible (A)-type counterpart to the (B)-view in the context of GR? For the purposes of this chapter, I take this to be the following: the metric field has a *primitive* connection to spacetime geometry, and in the regime in which terms featuring the Riemann tensor or its contractions may be ignored, the dynamical laws governing non-gravitational fields in a suitable neighbourhood of any point $p \in M$ are *constrained* to be invariant with respect to the local symmetries of this field in the same manner as for the (A)-story in the context of SR. Hence, the existence of the Lorentzian metric field *explains* the form of the local dynamical laws in the theory.[27] This view—which I will later, in Ch. 3, call an 'unqualified' geometrical view—thus embodies the target claim: that the metric field has its chronogeometric significance necessarily.

1.4.2 The Equivalence Principle and the Dynamical Perspective

Henceforth, I take (A) and (B) to refer to the above formulations of these views in the context of GR. How, then, do these fare in light of §1.3? As we have seen, the advocate of the (B)-view makes an inconsistent move insofar as she assumes the universal validity of **EP1** alongside minimal coupling. There is, however, a straightforward fix

[26]Strictly, the account given by the advocate of the dynamical perspective regarding the chronogeometricity of the metric field can go through at a point $p \in M$, so long as tidal forces may be ignored at that point (so that **EP1** holds approximately at that point). However, since (as discussed above) experimental apparatuses always occupy some extended region of spacetime, it is more natural to focus upon a neighbourhood of any $p \in M$, and thereby upon the approximate validity of **EP1'** for the fields measured by those apparatuses.

[27]I raise in §1.6 some worries regarding this account.

available: she may simply maintain that, to the extent that terms featuring the Riemann tensor or its contractions may be ignored, **EP1'** holds, and the metric field takes the form of the Minkowski metric in the relevant neighbourhood. These latter results suffice for the coherence of the dynamical approach in the context of GR.

The Poincaré invariance of local dynamical laws written in normal coordinates at any $p \in M$ merits further discussion. In particular, one might worry that there exists a tension between (I) the fact that such laws are invariant under *translations*, and (II) the fact that these laws generically do *not* hold at some other $q \in M$ in the neighbourhood of p (now setting aside the dropping of terms representing 'tidal forces'). To see this, suppose that one takes an active interpretation of such translations. In that case, one appears to find that the dynamical laws at p invariably *also* hold at q—a contradiction. Tension here can be resolved in the following way: one should *not* take an active interpretation of such translations upon the dynamical equations at p.[28] The reason for this is that, in writing the minimally coupled dynamical equations of GR at p in normal coordinates, information encoded in connection coefficients is not retained.[29] An actively-interpreted translation upon these equations at p would tell us that connection coefficients also vanish at q, *in the same coordinate system*—but we have prior knowledge that this is *not* the case. Hence, one should not interpret actively translations performed upon such dynamical equations, once written in normal coordinates at any $p \in M$.[30]

[28] Here, there are parallels with merely formal symmetries in the case of quantum mechanics, which admit of no active interpretation—see [25, §3.1].

[29] Note that this is due to the mathematical fact that connection components vanish at p in normal coordinates; no approximations are invoked in this discussion.

[30] This point is worth stressing, in order to avoid certain confusions. A metric field's being Minkowskian is a global rather than a local property of a manifold, for by definition all Minkowski metric fields are geodesically complete (cf. footnote 44). Thus, geometrically speaking, it is not clear how one can define 'local' Poincaré transformations, as discussed above—since from this point of view, such transformations are usually understood to be the isometries generated by the Killing vector fields on Minkowski spacetime. The resolution is to view the Poincaré transformations discussed above as passive only—their action is not on spacetime points at all, but rather on the chart space, i.e. the codomain of the coordinate charts.

1.5 Two Miracles of General Relativity

In response to claims to the effect that metric structure explains the symmetries of the equations governing non-gravitational fields, and in particular the fact that they are *all* Poincaré invariant, the advocate of (B) may reply that such an explanation is "question begging" [24, p. 139]. Regardless of which side one endorses in this debate, however, the fact that all such dynamical equations manifest the same symmetries is at least an *a priori* mystery; one might call it the *first miracle of relativity* (**MR1**):

MR1: All non-gravitational interactions are locally governed by Poincaré invariant dynamical laws.

MR1 holds in SR: it tells us that the dynamical laws governing all matter fields are Poincaré invariant; the advocate of the dynamical view (B) takes this to be a brute fact; the advocate of the geometrical view (A) attempts to rationalise this by appeal to Minkowski spacetime.[31] The fact that **MR1** holds has been discussed in [12, 24], and coheres well with the later thinking of Einstein: "The content of the restricted relativity theory can accordingly be summarized in one sentence: all natural laws must be so conditioned that they are covariant with respect to Lorentz transformations" [76, p. 329]. **MR1** still obtains in the neighbourhood of any $p \in M$ in GR. In addition to **MR1**, however, the work of §1.4 makes plausible that there exists in the GR context a *second miracle of relativity* (**MR2**):

MR2: The Poincaré symmetries of the dynamical laws governing non-gravitational

[31] On **MR1**, David Wallace has posed the following question: "Why should one think it a miracle that the intersection of the symmetry group of all the dynamical laws governing matter fields is the intersection of the symmetry group of all the dynamical laws governing matter fields?" To this, I respond as follows: the larger this intersection, the bigger the miracle. The fact that this intersection (and indeed also the union of the symmetry group of all dynamical laws governing matter fields) is the Poincaré group certainly qualifies, in my view, as 'miraculous', in the sense of footnote 3.

fields in the neighbourhood of any point $p \in M$ coincide—in the regime in which terms featuring the Riemann tensor or its contractions may be ignored—with the symmetries of the dynamical metric field in that neighbourhood.[32]

In GR, there exists an ontologically autonomous ('primordial') metric field, and this leads to **MR2**: why is it—assuming that terms representing tidal forces can be ignored—that the symmetries of the dynamical laws governing non-gravitational fields in a suitable neighbourhood of any $p \in M$ coincide with those of the primordial metric field in that neighbourhood? Again, the advocate of the dynamical view (B) may postulate this as a brute fact. By contrast, the advocate of the geometrical view (A)—in particular, the 'unqualified' geometrical view discussed in Ch. 3—may attempt to argue that the ontologically primitive metric field explains the form of the dynamical laws governing matter fields; however, as in SR, she faces an outstanding burden to delineate how this is supposed to work.[33] For more details on this, see §1.6, and Ch. 3.

One thing that the advocate of (A) may say here is the following: Minimally coupled dynamical laws in GR feature the metric field g_{ab}; as we have seen, the presence of (or rather, the coupling to) *this metric* constrains the local form in the neighbourhood of any $p \in M$ of the dynamical laws of those fields to which it couples. Consequently, the symmetries of the local dynamical laws *must* coincide with the symmetries of the metric field. This argument misses the point, however, for the very issue in question is *why* the dynamical laws governing matter fields take such a form—rather than another, with *different* local symmetry properties. In other words: *why this*

[32]Another way to put **MR2** would be the following: the signature of the metric field which codifies the local symmetries of the dynamical laws governing matter fields (cf. [30, §5]) coincides with the signature of the dynamical metric field appearing in Einstein's equation.

[33]Even supposing that an explanation can be found for **MR2**, there exist further questions in this vicinity. For example, why (with the programme of Callender in mind—cf. [34]) should the lobes of the lightcone structure designated as timelike by (a) the primordial metric field, and (b) the metric field which codifies the symmetries of the local dynamical laws, coincide?

particular coupling? This is the essence of **MR2**, which remains untouched by such arguments.

If one could argue that what had previously been regarded as the ontologically independent metric field of GR was in fact reducible to a codification of symmetries of the dynamical equations of matter fields (as with the (B)-account in SR), then this would provide an explanation for **MR2**. Indeed, one might think there are hints that this can be achieved: vacuum GR is essentially an *affine* theory, in the sense that the field equation of vacuum GR, $R_{ab} = 0$, only require an affine connection to be defined, as the Riemann and the Ricci tensor can be defined in terms of the connection only, without any need to refer to a metric tensor.[34] On the other hand, we *do* need a metric tensor to define the Ricci scalar that, together with the Ricci tensor, makes up the Einstein tensor that forms the left-hand side of the full Einstein equations.[35] Hence, one can go a long way in GR without defining a metric field—but as soon as one seeks to treat gravitational fields in the presence of matter, one must introduce a metric field.[36]

Note, however, that with only affine structure at one's disposal, there is no way to distinguish between four-dimensional space and four-dimensional space*time*. The reason is that one needs conformal or metric structure to distinguish between spatial and temporal dimensions, manifested in the signature of the metric (or of the conformal equivalence class of metrics). Thus, the claim that vacuum GR is essentially an affine theory needs to be treated with caution: we still need to obtain a metric or an equivalence class of metrics to make it a spacetime theory. Consequently, one

[34] I am grateful to Dennis Lehmkuhl for helpful discussions on these matters.

[35] The reason is that while the Ricci tensor can be obtained by contracting the one upper and one of the three lower indices of the Riemann tensor, in order to obtain the Ricci scalar we first need to raise one index of the Ricci tensor using the metric. This also means that we cannot define the standard Einstein-Hilbert Lagrangian without a metric.

[36] This coheres with the results of [121], that in order to define the stress-energy tensor of generic material systems, a metric tensor must be in place. We now find that both sides of the full Einstein equations require a metric in order to be well-defined.

must take even vacuum GR, understood as a theory of space*time*, to require metric structure, thereby raising potential obstacles for this move to explain **MR2**.[37]

Alternatively, one might seek to account for **MR2** by appeal to a particular form of the *relativity principle*, which states that the laws of physics take the same form in all inertial frames of reference. Recall that normal coordinates at any $p \in M$ are such that $g_{\mu\nu,\rho}(p) = 0$, and one may impose the further restriction that $g_{\mu\nu}(p) = \mathrm{diag}(-1, 1, 1, 1)$. It is plausible to regard such frames as *inertial*, since connection coefficients vanish therein; this means—via the geodesic equation—that motions of test bodies take their simplest form in such frames, insofar as they follow Newtonian inertial trajectories. Now consider a case in which non-gravitational fields governed by Galilean-invariant dynamical laws are coupled—via their associated stress-energy tensor—to Einstein's equation in which the curvature tensor is associated (via the unique torsion free, metric-compatible connection) to a Lorentzian metric field, g_{ab}. Suppose, for a given $p \in M$, that there exists a normal coordinate system in which the Galilean-invariant dynamical laws take their simplest form. Now transform to a new coordinate system, via a Lorentz boost. Though this new coordinate system is still normal—and so still inertial—the Galilean-invariant laws do *not* take the same form in this frame. Thus, this form of the relativity principle is violated. If one holds this principle to be sacrosanct, then one may, therefore, be able to account for **MR2**. Note, however, that this again merely pushes the problem back, for one may ask: *why* such a relativity principle?

[37] For presentations of GR in which affine structure is introduced before metric structure, see [128, 141, 194]—though note that, in such sources, this order of presentation is often chosen for pedagogical reasons, rather than to explore the extent to which it is possible to do GR with only affine structure, as is my concern here.

1.6 The Geometrical Perspective

The situation presented thus far stands as follows. (I): Minimal coupling implies **EP2**, but does not in general imply **EP1** (*pace* e.g. [24, ch. 9]). (II): When terms featuring the Riemann tensor or its contractions in local dynamical equations for non-gravitational fields can be ignored, **EP1'** holds in an appropriate neighbourhood of any $p \in M$. (III): Though philosophical positions such as the (B)-view are presented as being interwoven with claims regarding the compatibility of the universal validity of **EP1** and minimal coupling (see e.g. [24, ch. 9]), such claims may be excised from the position—for which suffices the approximate validity of **EP1'**. (IV): Some justification for restricting the allowed class of dynamical equations for non-gravitational fields in GR, or of the relativity principle, would suffice as an account of **MR1** and **MR2**.

By analogy with the case of SR, the advocate of the (A)-view (at least, what I will call in Ch. 3 the 'unqualified' (A)-view) is likely to claim that the sheer fact that non-gravitational fields are situated in a spacetime manifesting (under the appropriate conditions, i.e. when terms featuring the Riemann tensor or its contractions may be ignored) local Poincaré invariance in the neighbourhood of any $p \in M$ explains the local symmetries of their associated dynamical laws—and thereby explains **MR1** and **MR2**. As I see it, this is not correct. In this section, I present two interestingly distinct problem cases, in which the reasoning central to the (A)-view appears insufficient to account for the dynamical behaviour of matter fields. These two cases are the *Jacobson-Mattingly theory* on the one hand, and certain *bimetric theories* on the other (these theories are also discussed in [24, ch. 9]).[38]

[38]On my appeal to these theories, one might object: why is the fact that the (A)-view fails in such counterfactual scenarios, in which e.g. the Jacobson-Mattingly theory holds, of any relevance? The reason is that the (A)-view under consideration here postulates a strong modal constraint: all matter fields must be such that they 'advert' to all and only the designated metric structure, in the sense that the dynamical laws governing matter fields manifest (locally) the symmetries of that metric structure.

In the former (presented in e.g. [36, 106][39]), the action for a coupled Einstein-Maxwell system is augmented with an additional term (via a Lagrange multiplier field λ), imposing (as a field equation, via variation with respect to λ) that the vector potential A^a be locally timelike:[40]

$$S_{\mathrm{JM}}\left[g_{ab}, A^a, \lambda\right] = \int d^4x \sqrt{-g} \left(R - \frac{1}{4} F^{ab} F_{ab} + \lambda \left(g_{ab} A^a A^b - 1 \right) \right). \tag{1.6.1}$$

The imposition of this Lagrange multiplier term results in a violation of the relativity principle—though in a subtly different form to that presented in §1.5. The form of the relativity principle that is violated by the Jacobson-Mattingly theory is the following: dynamical laws for non-gravitational fields take *their simplest* form in all inertial frames. (Here, the 'simplest' form of a dynamical equation is to be understood as the form in which the greatest number of terms vanish—cf. appendix 1.B.) The reason that this version of the relativity principle is violated is that the Lagrange multiplier term in (1.6.1) picks out a preferred (timelike) direction in the neighbourhood of any $p \in M$; dynamical equations written in the subclass of inertial frames aligned with that direction admit of further simplification, in the sense that further terms vanish in those frames.[41] This notwithstanding, however, the form of the relativity principle

Scenarios such as those discussed in this section demonstrate that this need not be the case. Now, I leave open the possibility that there exist viable *refinements* of the (A)-view, which do not postulate such strong modal constraints—indeed, such possibilities are discussed in Ch. 3. In such cases, I concede that the force of the above cases may be mitigated. However, such developments of the (A)-view are *not* my concern in this chapter.

[39] In fact, the version of the Jacobson-Mattingly theory discussed in this chapter is a special case of that presented in [36,106].

[40] The first term is the Einstein-Hilbert action; F_{ab} is the Faraday tensor associated to A^a (see §1.2).

[41] For an illustration of how this term leads to the breaking of local Lorentz invariance—and so violations of this form of the relativity principle—see the discussion of a somewhat simpler (but analogous) theory in appendix 1.B. I do not discuss that simpler theory in the body of this chapter, for it is a theory set in fixed spacetime—whereas my concern in the above is (in part) to focus upon theories with dynamical metric structure, which manifest (violations of) **MR2**. Nevertheless, it is worth noting that the theory presented in appendix 1.B also raises questions regarding the kind of explanation for the

presented in §1.5 is still satisfied, for the dynamical equations of this theory still take the same Poincaré-invariant form in all inertial frames.

Note that the dynamical equations of the Jacobson-Mattingly theory *could* be constructed via minimal coupling;[42] this means that minimal coupling may lead to violations of certain forms of the relativity principle. In any case, however, the important point to make from the perspective of this section is that, in the Jacobson-Mattingly theory, the dynamical behaviour of non-gravitational fields does not reflect the local (Poincaré) symmetries of the metric field—taken to represent spacetime. The advocate of the (A)-view faces an outstanding burden to account for such cases.

Turn now to the second example: Bekenstein's bimetric T*e*V*e*S ('Tensor-Vector-Scalar') theory, presented in [15, 16]. As discussed in [24, §9.5.2], in this theory the metric field which is surveyed by rods and clocks, the conformal structure of which is traced by light rays, and the geodesics of which correspond to the motion of free bodies, is not the 'fundamental' metric field g_{ab}, but rather a less 'fundamental' metric field $\tilde{g}_{ab}$, constructed from the other fields in the theory [24, p. 174]. Thus, one might claim that the T*e*V*e*S theory presents another case in which the local symmetries of the dynamical laws do *not* mirror the local (Poincaré) symmetries of the 'background' metric field—a necessary condition for the (A)-view to go through.

Still, in this case, both g_{ab} and $\tilde{g}_{ab}$ are Lorentzian metric fields; moreover, the matter fields in this theory obey (in the relevant regime) locally Poincaré invariant dynamical laws. Thus, one might think that **MR2** is *satisfied* in this case, and thus that the example does not raise problems for the (A)-view. What is going on here? In fact, T*e*V*e*S points to an interesting ambiguity in **MR2**. If this principle states that the local

dynamical behaviour of matter proffered by advocates of the geometrical approach. As a second point on the Jacobson-Mattingly theory, the Lagrange multiplier term in the Jacobson-Mattingly action also leads to a violation of gauge invariance—cf. [106, p. 3].

[42] For example, the dynamical equation $g_{ab}A^aA^b = 1$ in the Jacobson-Mattingly theory could be constructed by applying the minimal coupling scheme to the associated dynamical equation $\eta_{ab}A^aA^b = 1$.

symmetry group of the dynamical laws governing matter fields is the same (in the relevant regime) as the local symmetry group of the metric field (in the sense given above), then it is indeed satisfied in T*e*V*e*S. Note, however, that in T*e*V*e*S it is the metric field $\tilde{g}_{ab}$, and not the metric field g_{ab}, which has chronogeometric significance—it is this field which takes a diagonal form in the frames in which the dynamical equations governing matter fields take their simplest form. If one reads **MR2** as demanding this stronger condition (i.e., if one takes a strong reading of the word 'coincides' in **MR2**), then T*e*V*e*S violates this principle for the 'fundamental' metric field g_{ab}, for the local symmetries of g_{ab} do not coincide in this strong sense with those of the dynamical laws governing matter fields. Since it is this stronger reading of **MR2** which is relevant to a field's having chronogeometric significance, it is this reading which should be preferred—in which case, T*e*V*e*S does pose a problem for the (A)-view, for it demonstrates that this local symmetry coincidence between the g_{ab} field and the dynamical equations governing matter fields does not hold of necessity.

1.7 Special Relativity

With all the above in mind, I now reflect on the nature of SR, and its relation to GR. Results such as (1.2.12) indicate that Poincaré invariance of dynamical laws in a spacetime theory is insufficient for that theory to be 'special relativistic', as such laws may still contain curvature pieces and derivatives of connection coefficients. Hence, some further criterion is needed to fully characterise 'special relativistic' theories. This criterion is that the inertial frames be *global*. To make sense of this, take the definition of an inertial frame presented in the previous section. As we have observed, $\Gamma^{\mu}{}_{\nu\rho}(p) = 0$ in normal coordinates at p (if the metric-compatible derivative operator is torsion-free), so this result holds in a local inertial frame. However, this is *only* generically true at p: at neighbouring points the connection coefficients may be non-zero, i.e. partial derivatives of the connection components may be non-zero in normal coordinates at p; this leads to the possibility of curvature terms in higher order equations in local inertial frames.

What happens, however, if we now specify that the inertial frames are *global*? In that case, connection components vanish globally in normal coordinates, so it follows that partial derivatives of the connection coefficients also vanish in these frames.[43] Hence, the Riemann curvature also vanishes globally. In such a case, higher order dynamical equations for non-gravitational fields in GR at any point $p \in M$ and written in normal coordinates take a Poincaré invariant form with no curvature terms, rather than merely a Poincaré invariant form simpliciter. Moreover, this result holds at *every* point in spacetime. Without specifying that the inertial frames are global, we do not recover such dynamical equations, either globally or (in general) locally. Hence, one might argue that 'special relativistic' theories are best characterised through two criteria:

1. The dynamical laws governing matter fields are Poincaré invariant.
2. The inertial frames are global.

Deciding whether to characterise such theories via (1) and (2) together, or just via (1), is a cost-benefit analysis. In the former case, we retain the notion that SR concerns Poincaré invariant dynamical laws governing matter fields in a fixed Minkowski spacetime.[44] However, in following this route we can no longer maintain that SR is in general recovered at a point in GR (although we can argue that it is *approximately*

[43] Note that I am not interested in the possibility of defining globally an arbitrary frame, but with the possibility that a given, globally-defined frame satisfies globally the condition of *inertiality*.

[44] Or at least, we almost do: in order to obtain Minkowski spacetime proper, we must specify that the manifold under consideration be diffeomorphic to $\mathbb{R}^4$, and geodesically complete.

recovered, in the regime in which **EP1'** holds). By contrast, in the latter case we *can* maintain that SR is locally recovered in GR. However, one has to greatly expand one's conception of the scope of SR to include the study of the dynamics of matter fields in generically non-flat spacetimes.

Though it might seem that the 'safe' option here is to endorse the former of the two approaches, on which 'special relativistic' theories are characterised by both (1) and (2) (especially as (2) is, arguably, an essential conceptual assumption of SR—recall that one of Einstein's great insights on his quest towards GR was precisely the *rejection* of (2)), it is at least worth countenancing the latter, more general account, according to which a theory's being 'special relativistic' is tied to the Poincaré invariance of its dynamical laws alone, and no further *a priori* restriction is made on the *content* of these laws. On this view, the assumption of a fixed, background Minkowski spacetime is *supererogatory* to the core of SR. Indeed, if one does choose to characterise special relativistic theories through (1) alone, then, given certain restrictions on the form of dynamical laws for matter fields (e.g., that they be constructed via minimal coupling), GR itself is a *locally* special relativistic theory.

1.8 Conclusions and Outlook

There exist circumstances in GR in which minimal coupling leads to violations of **EP1**. In this chapter, I have argued that this is unproblematic for advocates of the dynamical approach, for the relevant necessary condition for this view to account in the context of GR for the chronogeometricity of the dynamical metric field is, rather, the approximate validity of **EP1'**, which obtains in the regime in which terms featuring the Riemann tensor or its contractions in dynamical equations for non-gravitational fields may be ignored. As a result, this position remains a live option for accounting for chronogeometricity in GR, when dynamical equations for non-gravitational fields are selected in accordance with minimal coupling.

Exploring the above themes led me to identify **MR2**: that, when terms representing tidal forces may be ignored, the symmetries of the dynamical laws for non-gravitational fields in the neighbourhood of any point $p \in M$ coincide with the symmetries of the primordial metric field in that neighbourhood. It appears, much as with **MR1**, that—absent some deeper story—this must simply be treated as a brute fact in GR. Though an explanation of minimal coupling, or certain forms of the relativity principle, would suffice to account for **MR1** and **MR2**, no such explanation appears to be forthcoming.[45] Though the advocate of the (A)-view might claim to be able to account for both of these 'miracles', I have presented two cases which call into question such assertions.

Finally, I have reflected on the nature of SR. While one might choose to characterise 'special relativistic' theories via two conditions—(i) that dynamical laws are Poincaré invariant, and (ii) that inertial frames are global—in this chapter I have raised the possibility that such theories may be characterised solely by (i). Deciding which of these options is to be preferred is a nuanced business, worthy of future philosophical attention.

1.A Minimal Coupling and Poincaré Invariance

The goal of this appendix is to demonstrate that minimally coupled dynamical equations in GR manifest local Poincaré invariance, when written in normal coordinates at any $p \in M$. As a corollary, I derive some interesting results regarding when such

[45] At least within the domain of GR—it is possible that such an explanation can be found in e.g. perturbative string theory. See Ch. 2 for an exploration of this possibility.

equations exhibit a *broader* symmetry group. To begin, consider any minimally coupled dynamical equation in GR. In normal coordinates at some $p \in M$, this equation may schematically be written

$$O_{1,1} \cdots O_{1,n_1} + \ldots + O_{m,1} \cdots O_{m,n_m} = 0. \tag{1.A.1}$$

Here, m indexes the term; n_m indexes the object in the m-th term.[46] For such minimally coupled dynamical equations in normal coordinates, the set of relevant objects O_i featuring in such equations consists of (a) tensors; (b) partial derivatives of tensors; and (c) partial derivatives of connection components.[47] Given this, I now investigate the class of transformations under which (1.A.1) is invariant.[48] Applying an arbitrary coordinate transformation to (1.A.1), I write each O_i in the new, primed basis. Accordingly, the investigation reduces to an exploration of the individual transformation properties of each the O_i in such minimally coupled equations, in normal coordinates. There are three cases to consider: (a)-(c). Suppose first (a): that O_i is a tensor. Then, trivially, O_i transforms as a tensor under arbitrary coordinate transformations. So now suppose (b): that the O_i in question consists of a finite number of partial derivatives of a tensor. I claim that such an object transforms tensorially when the transformation in question is *affine*—i.e. when the transformation reads

$$x^{\mu} \to x^{\mu'} = M^{\mu'}{}_{\mu} x^{\mu} + a^{\mu'}, \tag{1.A.2}$$

[46]Indices in (1.A.1) are neither abstract nor coordinate indices, in the sense of the rest of this chapter.

[47]I do not include connection components in this enumeration, since these vanish in normal coordinates.

[48]A note on nomenclature. Consider an affine coordinate transformation $x^{\mu'} = M^{\mu'}{}_{\mu} x^{\mu} + a^{\mu'}$. If an (r,s) tensor field $T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}$ transforms under this coordinate change as $T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s} \to M^{\mu_1}{}_{\mu'_1} \ldots M^{\mu_r}{}_{\mu'_r} M^{\nu'_1}{}_{\nu_1} \ldots M^{\nu'_s}{}_{\nu_s} T^{\mu'_1 \cdots \mu'_r}{}_{\nu'_1 \ldots \nu'_s}$, then I say that it is *covariant* with this coordinate transformation. If, on the other hand, a dynamical equation retains the same form in either of the two coordinate systems under consideration, then I say that it is *invariant* under the coordinate change.

where the matrix elements ${M^{\mu'}}_{\mu}$ and vector components $a^{\mu'}$ are constant—in which case we have

$$\frac{\partial x^{\mu'}}{\partial x^{\mu}} = {M^{\mu'}}_{\mu} . \tag{1.A.3}$$

To show this, proceed by induction. Base case: O_i contains no partial derivatives of a tensor, so is just a tensor ($O_i = {T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s}$). Then O_i transforms tensorially under all coordinate transformations, *a fortiori* affine transformations. Inductive hypothesis: If O_i consists of n partial derivatives of a tensor, $O_i = \partial^{(n)}_{\sigma_n} \cdots \partial^{(1)}_{\sigma_1} {T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s}$, then $\partial^{(n-1)}_{\sigma_{(n-1)}} \cdots \partial^{(1)}_{\sigma_1} {T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s}$ transforms tensorially under affine coordinate transformations. Inductive step: For a given tensor ${T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s}$, if n partial derivatives of that tensor transforms tensorially under affine transformations, then show: so too does $O'_i = \partial^{(n+1)}_{\sigma_{n+1}} O_i$. Proof: By the inductive hypothesis, $O_i = \partial^{(n)}_{\sigma_n} \cdots \partial^{(1)}_{\sigma_1} {T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s}$ transforms tensorially under an affine coordinate transformation:

$$\begin{aligned} O_i &= \partial^{(n)}_{\sigma_n} \cdots \partial^{(1)}_{\sigma_1} {T^{\mu_1 \cdots \mu_r}}_{\nu_1 \ldots \nu_s} \\ &\to {M^{\sigma'_n}}_{\sigma_n} \cdots {M^{\sigma'_1}}_{\sigma_1} {M^{\mu_1}}_{\mu'_1} \cdots {M^{\mu_r}}_{\mu'_r} {M^{\nu'_1}}_{\nu_1} \cdots {M^{\nu'_s}}_{\nu_s} \partial^{(n)}_{\sigma'_n} \cdots \partial^{(1)}_{\sigma'_1} {T^{\mu'_1 \cdots \mu'_r}}_{\nu'_1 \ldots \nu'_s} , \end{aligned} \tag{1.A.4}$$

where the ${M^{\mu}}_{\mu'}$ are as per (1.A.2). Given this, O'_i transforms as

$$\begin{aligned} O'_i \to {M^{\sigma'_{n+1}}}_{\sigma_{n+1}} \partial^{(n+1)}_{\sigma'_{n+1}} \Big\{ & {M^{\sigma'_n}}_{\sigma_n} \cdots {M^{\sigma'_1}}_{\sigma_1} {M^{\mu_1}}_{\mu'_1} \cdots {M^{\mu_r}}_{\mu'_r} \cdot \\ & {M^{\nu'_1}}_{\nu_1} \cdots {M^{\nu'_s}}_{\nu_s} \partial^{(n)}_{\sigma'_n} \cdots \partial^{(1)}_{\sigma'_1} {T^{\mu'_1 \cdots \mu'_r}}_{\nu'_1 \ldots \nu'_s} \Big\} . \end{aligned} \tag{1.A.5}$$

Distributing the partial derivative using the product rule, all terms of the form $\partial^{(n+1)}_{\sigma_{n+1}} M^{\lambda'}{}_{\lambda}$ vanish, leaving

$$O'_i \to M^{\sigma'_{n+1}}{}_{\sigma_{n+1}} M^{\sigma'_n}{}_{\sigma_n} \cdots M^{\sigma'_1}{}_{\sigma_1} M^{\mu_1}{}_{\mu'_1} \cdots M^{\mu_r}{}_{\mu'_r} \cdot \\ M^{\nu'_1}{}_{\nu_1} \cdots M^{\nu'_s}{}_{\nu_s} \partial^{(n+1)}_{\sigma'_{n+1}} \partial^{(n)}_{\sigma'_n} \cdots \partial^{(1)}_{\sigma'_1} T^{\mu'_1 \ldots \mu'_r}{}_{\nu'_1 \ldots \nu'_s}. \quad (1.A.6)$$

Hence, if the object O_i consisting of n partial derivatives of a tensor $T^{\mu_1 \ldots \mu_r}{}_{\nu_1 \ldots \nu_s}$ is invariant under affine coordinate transformations, then so too is the object O'_i consisting of $(n+1)$ partial derivatives of that same tensor. Thus, all partial derivatives of tensors appearing in minimally coupled dynamical equations at $p \in M$ transform tensorially under affine coordinate transformations.

Finally, consider (c). The claim now is that all partial derivatives of connection components transform tensorially under affine coordinate transformations. To prove this, again proceed by induction. Base case: $O_i = \partial_\sigma \Gamma^{\mu}{}_{\nu\lambda}$, i.e. O_i consists of one partial derivative of a connection component. Transforming to a new coordinate basis and then expanding in terms of the old basis, we have

$$\begin{aligned}
\partial_{\rho'} \Gamma^{\kappa'}{}_{\sigma'\nu'} &= \frac{\partial x^{\rho}}{\partial x^{\rho'}} \frac{\partial}{\partial x^{\rho}} \left(\frac{\partial x^{\sigma}}{\partial x^{\sigma'}} \frac{\partial x^{\nu}}{\partial x^{\nu'}} \Gamma^{\kappa}{}_{\sigma\nu} \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} + \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} \frac{\partial^2 x^{\kappa}}{\partial x^{\sigma'} \partial x^{\nu'}} \right) \\
&= \frac{\partial^2 x^{\sigma}}{\partial x^{\rho'} \partial x^{\sigma'}} \frac{\partial x^{\nu}}{\partial x^{\nu'}} \Gamma^{\kappa}{}_{\sigma\nu} \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} + \frac{\partial x^{\sigma}}{\partial x^{\sigma'}} \frac{\partial^2 x^{\nu}}{\partial x^{\rho'} \partial x^{\nu'}} \Gamma^{\kappa}{}_{\sigma\nu} \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} \\
&\quad + \frac{\partial x^{\rho}}{\partial x^{\rho'}} \frac{\partial x^{\sigma}}{\partial x^{\sigma'}} \frac{\partial x^{\nu}}{\partial x^{\nu'}} \frac{\partial}{\partial x^{\rho}} \Gamma^{\kappa}{}_{\sigma\nu} \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} + \frac{\partial x^{\rho}}{\partial x^{\rho'}} \frac{\partial x^{\sigma}}{\partial x^{\sigma'}} \frac{\partial x^{\nu}}{\partial x^{\nu'}} \Gamma^{\kappa}{}_{\sigma\nu} \frac{\partial^2 x^{\kappa'}}{\partial x^{\rho} \partial x^{\kappa}} \\
&\quad + \frac{\partial x^{\rho}}{\partial x^{\rho'}} \frac{\partial^2 x^{\kappa'}}{\partial x^{\rho} \partial x^{\kappa}} \frac{\partial^2 x^{\kappa}}{\partial x^{\sigma'} x^{\nu'}} + \frac{\partial x^{\rho}}{\partial x^{\rho'}} \frac{\partial x^{\kappa'}}{\partial x^{\kappa}} \frac{\partial}{\partial x^{\rho}} \frac{\partial^2 x^{\kappa}}{\partial x^{\sigma'} \partial x^{\nu'}}.
\end{aligned} \quad (1.A.7)$$

Five of the six terms in (1.A.7) contain second partial derivative pieces, which vanish when the coordinate transformation is affine. So, under an affine coordinate transformation, this object transforms as a tensor. Inductive hypothesis: If O_i consists of n partial derivatives of a connection component, $O_i = \partial^{(n)}_{\sigma_n} \cdots \partial^{(1)}_{\sigma_1} \Gamma^{\mu}{}_{\nu\lambda}$, then $\partial^{(n-1)}_{\sigma_{(n-1)}} \cdots \partial^{(1)}_{\sigma_1} \Gamma^{\mu}{}_{\nu\lambda}$ transforms tensorially under affine coordinate transformations. Inductive step: If O_i consists of n partial derivatives of a connection component and transforms tensorially under affine transformations, then show: so too does $O'_i = \partial^{(n+1)}_{\sigma_n+1} O_i$. Proof: *mutatis mutandis* as for the inductive step of (b). Thus, all partial derivatives of connection components appearing in minimally coupled dynamical equations at $p \in M$ transform tensorially under affine coordinate transformations.

We have found that each of the O_i featuring in any minimally coupled dynamical equation in GR, written in normal coordinates at a point $p \in M$, is covariant—i.e., transforms tensorially—under affine coordinate transformations. However, we have yet to show that all such equations are *invariant*—i.e. take the same form—under affine coordinate transformations. In fact, this is in general *not* the case: and whether equations of the form (1.A.1) are so invariant, or rather are invariant under a *restricted* class of affine transformations (most relevantly the Poincaré group) depends upon context. To see this, first consider (1.2.9). Written in normal coordinates at some $p \in M$, the equation reads

$$F_{[\mu\nu,\lambda]} = 0. \tag{1.A.8}$$

Transforming to a new coordinate basis under an affine transformation, (1.A.8) reads, using the above results,

$$M^{\mu'}{}_{\mu} M^{\nu'}{}_{\nu} M^{\lambda'}{}_{\lambda} F_{[\mu'\nu',\lambda']} = 0. \tag{1.A.9}$$

At this stage, however, assuming that the transformation matrices are invertible, we may simply operate on (1.A.9) with the relevant matrix inverses, yielding

$$F_{[\mu'\nu',\lambda']} = 0. \tag{1.A.10}$$

Hence, (1.2.9) written in normal coordinates at any $p \in M$ is invariant under all affine transformations with invertible linear transformation matrices. Now, however, consider (1.2.8). In normal coordinates at some $p \in M$, this reads

$$F^{\mu\nu}{}_{,\nu} = J^{\mu}. \tag{1.A.11}$$

Transforming to a new coordinate basis under an affine transformation, (1.A.11) becomes

$$M^{\mu}{}_{\mu'} M^{\nu}{}_{\nu'} M_{\nu}{}^{\lambda'} F^{\mu'\nu'}{}_{,\lambda'} = M^{\mu}{}_{\mu'} M^{\nu}{}_{\nu'} M^{\sigma}{}_{\lambda'} \eta_{\nu\sigma} F^{\mu'\nu',\lambda'} = M^{\mu}{}_{\mu'} J^{\mu'}. \tag{1.A.12}$$

While the $M^{\mu}{}_{\mu'}$ matrices can be cancelled (via multiplication by the relevant inverse matrices) in (1.A.12), in order for this equation to be *invariant* under the affine coordinate transformation in question, we require the further condition that

$$M^{\nu}{}_{\nu'} M^{\sigma}{}_{\lambda'} \eta_{\nu\sigma} = \eta_{\nu'\lambda'}. \tag{1.A.13}$$

From this, we see that (1.2.8), written at some $p \in M$ in normal coordinates, is invariant *only* under the more restricted class of *Poincaré transformations*—since (1.A.13) is the definition of a Lorentz rotation. In general, whether the minimally coupled dynamical equations of matter fields in GR, written at a point in normal coordinates, are invariant under all affine transformations (with invertible linear transformation matrices), or only the Poincaré transformations, depends upon whether the metric field is used to contract indices between the O_i. To summarise:

- Objects of the form (a)-(c) transform tensorially under affine transformations; since the dynamical equations of GR obtained via minimal coupling, written at a point in normal coordinates, contain only objects of the form (a)-(c), all objects in such equations are covariant under affine transformations.

- Due to the potential contraction of indices in some terms of such equations with respect to the metric, we sometimes require further conditions on the affine transformations in question for that equation to be *invariant* thereunder. Where the metric is the Minkowski metric, this condition is that the transformations be Poincaré transformations.

The latter point is important. To illustrate, consider (1.2.1). We do not know under *which* affine transformations this equation is invariant until we specify the metric field with respect to which index contraction takes place. If this is the Minkowski metric, then such transformations are the Poincaré transformations, by the above reasoning.

If, however, this is some other metric, then the affine transformations under which such objects are invariant will differ.[49]

1.B Lorentz Symmetry Breaking

Following [35, pp. 1231ff.], consider the Lagrangian for electrodynamics set in a fixed Minkowski spacetime as per §1.2.1, but now augmented with a Chern-Simons term,[50]

$$\mathcal{L}_{\mathrm{CFJ}} = \mathcal{L}_{\mathrm{EM}} - \frac{1}{2} p_a A_b \tilde{F}^{ab}. \tag{1.B.1}$$

Here, $\tilde{F}^{ab} := \frac{1}{2}\epsilon^{abcd}F_{cd}$ is the dual electromagnetic tensor. This modification couples electromagnetism to an (as yet unspecified) vector field, p^a.[51] At the level of kinematically possible models, p^a is specified to be *spacelike*, so that $p_a p^a \equiv m^2 > 0$. Varying the action associated to (1.B.1), one obtains a generalisation of (1.2.1),

$$F^{ab}{}_{;b} = J^a + p_b \tilde{F}^{ba}; \tag{1.B.2}$$

by contrast, (1.2.2) still holds. (Note that $p_a := \eta_{ab} p^b$.) Given that p^a is spacelike, one

[49] Cf. [94, 237]. As noted in [93, 105], the specific spacetime metric implicit in one's theory of electrodynamics is specified by the so-called *constitutive relations*; for a certain simple form of the constitutive relations involving the *Hodge dual operator*, this is the Minkowski metric.

[50] For background on Chern-Simons theory, see e.g. [62]. Another theory which could be used to illustrate the same point is SR with a homogeneous dust field, the four-velocity field of which generates a timelike geodesic congruence.

[51] As with the Minkowski metric field η_{ab} in the version of electromagnetism introduced in §1.2.1, the vector field p^a may be understood to be a *fixed field*, in the sense of [171, p. 13].

can select a frame in which $p^0 = 0$, in which case (1.B.2) reduces to

$$F^{\nu\mu}{}_{;\mu} = J^\nu + p_i \tilde{F}^{i\nu}. \tag{1.B.3}$$

Together, (1.B.3) and (1.2.2)[52] are the simplest forms of the dynamical equations for this theory. The question, then, is: under which coordinate transformations are the forms of these equations preserved? Recalling from appendix 1.A that (1.2.1) is invariant under Poincaré transformations, while (1.2.2) is invariant under all affine transformations with invertible linear transformation matrix, it is clear that the forms of (1.B.3) and (1.2.2) will together be invariant under the subgroup of Poincaré transformations that preserve the condition $p^0 = 0$. Our task now, then, is to identify this class of transformations.

By acting upon p^a with an arbitrary transformation matrix, one finds that the coordinate transformations which preserve $p^0 = 0$ are the Galilean transformations.[53] Thus, the class of transformations which preserve (1.B.3) and (1.2.2) together consists of the intersection of the Poincaré and Galilean transformations—these are the translations and spatial rotations. Thus, in this modified theory of electrodynamics, the presence of the p^a vector field 'breaks' Lorentz symmetry, reducing the class of transformations under which the dynamical equations of the theory hold in their simplest form from the Poincaré transformations to the translations and spatial rotations. This is reconcilable with the Poincaré invariance of (1.B.2) and (1.2.2) together, since in

[52]The latter written in a coordinate basis.

[53]To see this, first act with a generic linear transformation matrix $M^\mu{}_\nu$ (associated to an affine coordinate transformation $x'^\mu = M^\mu{}_\nu x^\nu + a^\mu$—here I have moved the primes from indices to the new coordinates, i.e. have written x'^μ rather than $x^{\mu'}$, for clarity of exposition) upon p^μ (this geometrical object now being written in some coordinate basis), subject to the condition that the temporal component also vanishes in the new coordinate system, i.e. $p'^0 = 0$. Doing so, one finds that the time-space component $M^0{}_i$ of the transformation matrix vanishes. Given this, it follows that $x'^0 = M^0{}_0 x^0 + a^0$ and $x'^i = M^i{}_0 x^0 + M^i{}_j x^j + a^i$; these are precisely the Galilean transformations.

this case these do *not* constitute the simplest forms of the dynamical equations of the theory. A similar result holds for the *local* dynamical equations at any $p \in M$ for the general relativistic Jacobson-Mattingly theory [36, 106], discussed in §1.6.

Chapter 2

On Miracles and Spacetime

What were dubbed in the previous chapter the two 'miracles' of general relativity remain, I have argued, unaccounted for in that theory. In this chapter, I demonstrate that these two 'miracles' admit of a natural explanation in one particular successor theory to general relativity—namely, perturbative string theory. I argue that this point has important implications when considering both the 'chronogeometricity' (that is, the object in question being surveyed by rods and clocks built from matter fields) and spatiotemporal status of the dynamical metric field in both general relativity and perturbative string theory.

2.1 Introduction

In the previous chapter, I drew attention to two 'miracles'—i.e., features surprising or puzzling, the explanation of which remains outstanding—in the foundations of general relativity (GR).[1] These 'miracles', labelled respectively **MR1** and **MR2**, can be stated as follows:

MR1: All non-gravitational interactions are locally governed by Poincaré invariant dynamical laws.

MR2: The Poincaré symmetries of the dynamical laws governing non-gravitational fields in the neighbourhood of any point in the manifold coincide—in the regime in which terms featuring the Riemann tensor or its contractions can be ignored—with the symmetries of the dynamical metric field in that neighbourhood.

In this chapter, after briefly recalling the details of **MR1** and **MR2**, I demonstrate that while these two results are indeed 'miraculous' in GR, they cease to be so in one particular successor theory of quantum gravity to GR, namely perturbative string theory. Not only is this a noteworthy observation in itself, but (I contend) it has important consequences when assessing both the 'chronogeometricity' (that is, the object in question being surveyed by rods and clocks built from matter fields[2]) and spatiotemporal status of the dynamical metric field in both general relativity and perturbative string theory.

To see the sense in which this is so, note two points. First, **MR1** and **MR2** constitute important conditions for the dynamical metric field in GR to acquire what is referred to in [24, 29, 180] as its 'chronogeometric significance'—that is, for it to be surveyed by rods and clocks built from matter fields (see §2.3). In brief, the reason

[1] To repeat: the reader is urged not to be distracted by the nomenclature of 'miracles'; cashing out exactly what qualifies as a 'miracle' is the central concern of neither the previous nor the present chapter. Rather, the above-quoted characterisation is fully sufficient for our purposes.

[2] Recall that, roughly speaking, by 'surveyed by rods and clocks built from matter fields', I mean that rods and clocks built from matter fields can be used to read off intervals of the metric field; this notion is drawn from the work of Brown (see e.g. [24, ch. 9]), and is discussed in more detail in §2.3 below.

for this is that **MR2** ensures a surveying of the metric field by rods and clocks; **MR1** ensures the irrelevance of the constituents of rods and clocks.[3] Second, if one follows Knox in defining 'spacetime' *functionally*, so that "the spacetime role is played by whatever defines a structure of local inertial frames" [113, p. 9], then, as we shall see, the obtaining of **MR1** and **MR2** is also an important condition for the metric field in GR to qualify as *spatiotemporal*. The fact that **MR1** and **MR2** are *contingent* in GR means that both chronogeometricity and spatiotemporality in GR is also contingent. Not so in perturbative string theory, where these two 'miracles' follow ineluctably from the formalism of the theory.

Consequently, both the chronogeometricity and spatiotemporality of the (non-fundamental) dynamical metric field in perturbative string theory have a stronger status than in GR. This prompts me to distinguish (A) the objects which play the functional role of spacetime *à la* Knox being fundamental in the theory, from (B) the objects which play the functional role of spacetime doing so necessarily rather than contingently. While GR exemplifies (A) but not (B), the situation in perturbative string theory is the reverse: the theory exemplifies (B), but not (A).

The structure of this chapter is as follows. In §2.2, I discuss **MR1** and **MR2**, and why these 'miracles' are merely contingent in GR. In §2.3, I discuss the chronogeometric and (for Knox) spatiotemporal status of the metric field in GR. In §2.4, I recall the relevant details of perturbative string theory, and demonstrate that, in this theory, **MR1** and **MR2** are necessary rather than contingent, and admit of a natural explanation in the string-theoretic context—thereby rendering them miraculous no more. In §2.5, I say a few more words on the logic underlying these 'miracles'. In §2.6, I draw upon these results in order to compare the spatiotemporal status of the dynam-

[3]Note, though, that **MR1** and **MR2** do not in themselves constitute sufficient conditions for such chronogeometricity, for one must also presuppose e.g. the existence of stable rods and clocks built from matter fields.

ical metric field in perturbative string theory, versus in GR. In §2.7, I discuss how Huggett's arguments in [100] that 'target space' in perturbative string theory is not 'phenomenal space' connect with these matters.

2.2 Two Miracles

In this section, I recall the content of **MR1** (§2.2.1) and **MR2** (§2.2.2) and their contingent status in GR, as a prelude to demonstrating in §2.4 how this situation changes on moving to perturbative string theory.

2.2.1 The First Miracle: Universal Local Poincaré Invariance

In order to understand **MR1**, one must recall some details regarding the dynamical equations of GR. The kinematically possible models (KPMs) of GR are picked out by triples $\langle M, g_{ab}, \Phi \rangle$, where M is the a four-dimensional differentiable manifold; g_{ab} is a generic Lorenzian metric field on M; and Φ is a placeholder for matter fields in the theory. Dynamically possible models (DPMs) of GR are those KPMs the geometrical objects of which obey certain specified dynamical equations: the Einstein equation,

$$G_{ab} := R_{ab} - \frac{1}{2} g_{ab} R = 8\pi T_{ab}, \tag{2.2.1}$$

i.e. the dynamical equation associated with g_{ab}, *plus* dynamical equations for the matter fields.[4] (Here, T_{ab} is the stress-energy tensor associated with the matter fields Φ.)

[4] Strictly, independence of these dynamical equations from the Einstein equation depends on the case in question—for discussion of this point, see [24, §9.3] and [141, §20.6].

Ab initio, there is no restriction on the possible dynamical equations for matter fields in GR. However, one often restricts consideration to equations which are *minimally coupled*. Recall from Ch. 1 that the prescription for constructing such minimally coupled equations is the following:

> Minimally coupled dynamical equations for matter fields in GR are constructed from dynamical equations for matter fields featuring coupling to a fixed Minkowski metric field η_{ab} and no curvature terms, by replacing all instances of η_{ab} with a generic Lorentzian metric field g_{ab}, and replacing all instances of the torsion-free derivative operator compatible with η_{ab} with the torsion-free derivative operator compatible with g_{ab}.

Sometimes, this is referred to as the 'comma-to-semicolon' prescription. To make further progress in understanding minimal coupling, recall again from Ch. 1 some basic facts regarding pseudo-Riemannian geometry. In a coordinate basis $\{e_\mu\}$, the *connection components* $\Gamma^\mu{}_{\nu\rho}$ associated to a derivative operator ∇_a are defined by $\nabla_\rho e_\nu =: \Gamma^\mu{}_{\nu\rho} e_\mu$. Then, at any $p \in M$, we can choose *normal coordinates*, such that $\Gamma^\mu{}_{(\nu\rho)}(p) = 0$ in those coordinates; for a torsion-free derivative operator, we can in fact choose normal coordinates such that $\Gamma^\mu{}_{\nu\rho}(p) = 0$. (Note that the connection components away from p will in general *not* vanish.) If the unique torsion-free, metric compatible derivative operator is used, then in normal coordinates we also have $\partial_\rho g_{\mu\nu}(p) = 0$, and we may further restrict to the subclass of normal coordinates at p such that $g_{\mu\nu}(p) = \mathrm{diag}(-1, 1, 1, 1)$.

Minimally coupled dynamical equations for non-gravitational fields in GR take a Poincaré invariant form in normal coordinates at any $p \in M$—see Ch. 1. Indeed, the work of that chapter demonstrates that *any* dynamical equation featuring coupling to (a) g_{ab}; (b) its associated curvature tensor; (c) the matter fields in the theory, will also take Poincaré invariant form in normal coordinates at any $p \in M$—even if such a dynamical equation is not obtained via the minimal coupling prescription. On the other hand, many dynamical equations are possible in GR which are *not* locally Poincaré invariant—for example, there is nothing to rule out dynamical equations for non-gravitational fields in GR which are locally *Galilean* invariant.[5]

The question posed by **MR1**, then, is: why should the dynamical equations for non-gravitational fields be such that all are Poincaré invariant? Given that other coupling schemes are possible in which such local Poincaré invariance does not obtain, that such *is* the case constitutes a 'free lunch'. Within GR, it does not appear that this result can be delivered without further input assumptions—e.g., the restriction to minimally coupled dynamical equations. But in any case, satisfaction of **MR1** must ultimately simply be *postulated*; it does not appear that any deeper explanation of this 'miracle' is possible, within the framework of GR.[6]

2.2.2 The Second Miracle: Metric Symmetry Coincidence

Turn now to an explication of **MR2**. First, recall that, in normal coordinates at p, $\partial_\rho g_{\mu\nu}(p) = 0$, and one may further restrict to the subclass of normal coordinates such that $g_{\mu\nu}(p) = \mathrm{diag}(-1,1,1,1)$; however, higher derivatives of the metric field do not necessarily vanish at p. Second, assume that we are dealing with matter fields which

[5]Presumably, KPMs here would be picked out by tuples $\langle M, g_{ab}, t_{ab}, h^{ab}, \nabla_a, \Phi\rangle$, where g_{ab} is a generic Lorentzian metric field satisfying (2.2.1); $\langle M, t_{ab}, h^{ab}, \nabla_a\rangle$ is the structure of 'classical spacetime' (in the sense of [128, ch. 4]) necessary to write down Galilean invariant laws for non-gravitational fields; Φ are the matter fields, which obey Galilean invariant dynamical laws; and the stress-energy tensor T_{ab} in (2.2.1) is that associated with the Φ. In this scenario, g_{ab} dynamically evolves in response to the stress-energy content of the Φ, but the Φ are not so sensitive to g_{ab}—for this metric field does not feature in their associated dynamical equations. For this reason, we seem to have here instantiation of a strange, *reverse* violation of the action-reaction principle, since the metric field is responsive to matter fields but not vice versa (for further discussion of the action-reaction principle, see [24, 26]).

[6]For further (tentative) reflections on this point, related to the so-called *geodesic principle* in GR, see Ch. 3.

do all satisfy Poincaré invariant dynamical laws at any $p \in M$. Third, note that such dynamical equations for matter fields may still feature curvature terms, even at p (for an extended discussion of this point, see Ch. 1). Since higher derivatives of the metric field do not vanish at p, it would be misleading to state that the metric field g_{ab} 'reduces' to the Minkowski metric field η_{ab} at p. However, since non-vanishing higher derivatives of the metric field are associated with non-vanishing derivatives of connection components, which in turn are associated with non-vanishing curvature, we can say that, *in the regime in which terms featuring derivatives of connection components can be neglected* (whether this is so depends upon the strength of one's experimental apparatus relative to the strength of the 'tidal gravitational forces' represented by these terms—for detailed discussion, see Ch. 1), the metric field g_{ab} *does* recover the form of a Minkowski metric η_{ab} at p, *and* the dynamical laws governing matter fields reduce to a Poincaré invariant form with no curvature terms, at p.

What this means is that, in such a regime, the symmetries of the metric field at $p \in M$—i.e., the coordinate transformations under which the metric at p retains the same form—coincide with the symmetries of the dynamical laws governing matter fields at p—i.e., the coordinate transformations under which these dynamical equations retain the same form. That such *is* the case is a statement of **MR2**—again, this constitutes a contingent 'miracle' in GR, for it is in principle possible that all matter fields are governed by dynamical equations with the same local symmetries, which nevertheless do not coincide with those local symmetries of the metric field—the above example of all matter fields obeying Galilean invariant local dynamical laws constitutes just such a case.

2.3 Chronogeometry and Spacetime

Why is it that intervals of the metric field g_{ab} in GR can be read off from rods and clocks constructed from matter fields? An important condition for the metric field of GR to have 'chronogeometric significance' in this sense is that **MR1** and **MR2** hold. To see this, suppose that these principles hold, so that all dynamical laws for matter fields have the same symmetries, which locally coincide (in the appropriate regime) with those of the metric field. Now consider a stable rod or clock built from matter fields, in a particular frame.[7] If such a rod or clock is a 'good' rod/clock in that frame (if, one might say, it is 'ideal' in that frame), then it can be used to measure intervals as given by the metric field, in that frame. But now, since **MR1** and **MR2** are assumed to hold, in a Poincaré-transformed frame such is *still* the case, for metric and dynamical equations for matter fields transform in the same way—meaning that this stable rod/clock yields a robust means of measuring such metric intervals in all frames related by Poincaré transformations.[8]

Since **MR1** and **MR2** are contingent in GR, so too is this chronogeometric significance of the metric field: if different matter fields obeyed different dynamical laws, then the metric field would not have (universal) local chronogeometric significance; if the symmetries of dynamical laws for matter fields did not coincide locally with those of the metric field, then, again, the metric field would not have local chronogeometric significance. Indeed, not only is the chronogeometric significance of the metric field in GR a contingent matter, but so too is its *spatiotemporal status*, on a spacetime functionalist approach such as that of Knox [111–113]. Recall again Knox's slogan: "I propose that the spacetime role is played by whatever defines a structure of local

[7]Recall from footnote 3 that the existence of such stable rods and clocks constitutes an important additional condition for the metric field of GR to have chronogeometric significance.

[8]By contrast, suppose that one has a rod/clock built from matter fields which obey Galilean invariant laws; then, in a Poincaré-transformed frame, that rod/clock would in general not obey dynamical equations of the same form—and so would in general not be a 'good' rod/clock in that frame.

inertial frames” [113, p. 9]. The idea is to functionally *define* spacetime as any structure which itself picks out a structure of local inertial frames. In turn, Knox gives the following characterisation of inertial frames: [111, p. 348]

> In Newtonian theories, and in special relativity, inertial frames have at least the following three features:
>
> 1. Inertial frames are frames with respect to which force free bodies move with constant velocities.
> 2. The laws of physics take the same form (a particularly simple one) in all inertial frames.
> 3. All bodies and physical laws pick out the same equivalence class of inertial frames (universality).

Any structure which picks out a “structure of local inertial frames”, i.e. a structure of local frames which satisfy these properties (initially identified as significant in the Newtonian/special relativistic context) qualifies definitionally as ‘spacetime’, for Knox. But note that, in GR, **MR1** and **MR2** precisely guarantee that this field be considered spatiotemporal, in this sense. The reason is that, locally, the symmetries of the dynamical metric field coincide with those of the dynamical equations governing matter fields; in any frame in which these dynamical equations take their simplest form, the metric field itself takes the form $\mathrm{diag}\,(-1,1,1,1)$. Thus, the metric field picks out a structure of local inertial frames.

2.4 Perturbative String Theory

The above in hand, I turn now to perturbative string theory—one candidate quantum theory of gravity.[9] Classically, a string can be regarded as a special case of a p-brane, which is an object with p dimensions and tension $T_p = 1/(2\pi\alpha')$, where α' is the 'Regge slope parameter'.[10] The classical motion of a p-brane extremises the $(p+1)$-dimensional volume V that it sweeps out in a D-dimensional 'target space'.[11] Thus, there is a p-brane action given by $S_p = -T_p V$. In the case of the fundamental string, which has $p = 1$, V is the area of the string worldsheet and the action is called the *Nambu-Goto action* [22, p. 10]

$$S_{\text{NG}} = -\frac{1}{2\pi\alpha'}\int_{\Sigma} d^2\sigma \left[-\det_{\mathfrak{ab}}\left(\frac{\partial X^a}{\partial\sigma^{\mathfrak{a}}}\frac{\partial X^b}{\partial\sigma^{\mathfrak{b}}}\eta_{ab}\right)\right]^{1/2}, \tag{2.4.1}$$

where Σ denotes the string worldsheet; the functions $X^a(\sigma,\tau)$ describe the target space embedding of the worldsheet; $\tau \equiv \sigma^0$ and $\sigma \equiv \sigma^1$ are coordinates on the worldsheet (the parameter τ is the worldsheet 'time' coordinate, while σ parametrises the string at a given worldsheet 'time'); $d^2\sigma = d\tau d\sigma$; and Fraktur script denotes worldsheet indices. Classically, the Nambu-Goto action is equivalent to the string sigma-

[9] String theory goes further than other theories of quantum gravity (e.g. loop quantum gravity), since it purports to realise the more ambitious goal of unifying all four fundamental forces into one (quantum mechanical) framework. That is, string theory purports to be a 'theory of everything', whereas other quantum gravity theories need not do so. For more details on theories of quantum gravity, see Ch. 5.

[10] The Regge slope parameter is equal to the square of the fundamental string length—see e.g. [14, 165] for details.

[11] Recently, in light of so-called *dualities* in string theory (see e.g. [184], and Chs. 5 and 6), Huggett has argued that "phenomenal spacetime" is *not* equivalent to target space [100]—see §2.7 for an extended discussion of this point.

model action, also known as the *Polyakov action* [22, p. 13]

$$S_{\mathrm{P}(\eta)} = -\frac{1}{4\pi\alpha'}\int_{\Sigma} d^2\sigma\sqrt{-h}h^{\mathfrak{ab}}\eta_{ab}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b, \tag{2.4.2}$$

where $h_{\mathfrak{ab}}(\tau,\sigma)$ is an auxiliary[12] worldsheet metric with inverse $h^{\mathfrak{ab}}(\tau,\sigma)$; and $h := \det h_{\mathfrak{ab}}$. The Euler-Lagrange equation for $h_{\mathfrak{ab}}$ can be used to eliminate it from (2.4.2) and recover (2.4.1).[13] Quantum mechanically, instead of eliminating $h_{\mathfrak{ab}}$ via its classical field equations, one should perform a path integral, using standard machinery to deal with the local symmetries and gauge fixing.[14] Doing this, one finds that there is a conformal anomaly[15] unless the target space dimension is $D = 26$.[16]

For a closed string, one imposes periodicity in σ. Choosing its range to be π, one identifies both ends of the string $X^a(\tau,\sigma) = X^a(\tau,\sigma+\pi)$. After quantising and defining suitable ladder operators (see e.g. [14, p. 53]), one can act on the ground state of the string with raising operators to study its spectrum. For the closed string, there exist three distinct first excited states, denoted f_{ab}, B_{ab}, and Φ. f_{ab}—suggestively christened the *graviton*—is symmetric and traceless in its target space indices, and

[12]In the sense that $h_{\mathfrak{ab}}$ is a *new* variable, *a priori* independent of the pullback of the target space metric to the worldsheet.

[13]Varying (2.4.2) with respect to $h^{\mathfrak{ab}}$, we obtain the equation of motion $\eta_{ab}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b - \frac{1}{2}h_{\mathfrak{ab}}h^{\mathfrak{cd}}\eta_{ab}\partial_{\mathfrak{c}}X^a\partial_{\mathfrak{d}}X^b = 0$; back-substitution then returns (2.4.1).

[14]I.e. path integral quantisation *à la* Faddeev-Popov—see e.g. [22, §3.4].

[15]An *anomaly* arises when a symmetry of a classical theory is not manifested in the associated quantum theory. In more technical language, though the classical action is invariant under the symmetry, the associated path integral measure—used to define the quantum theory—is not. The *conformal anomaly* arises on quantisation of classical string theory, and breaks the conformal invariance of the string worldsheet. This anomaly manifests itself as an extra term in the *Virasoro algebra*, which comprises the generators of the conformal group of the string worldsheet; this extra term vanishes—thereby circumventing the anomaly—only in the case of target space dimension $D = 26$ for the bosonic string, or $D = 10$ for the fermionic string. For a philosophically-oriented introduction to anomalies, see [99, §4].

[16]As mentioned in footnote 15, analogous analysis for superstrings (i.e. strings for which supersymmetry is added—either on the worldsheet as in the so-called *RNS sector*, or to the background target space as in the *GS sector*)—gives the critical dimension $D = 10$ [14, p. 7]. In this chapter I focus solely on bosonic string theory.

transforms under $SO(D-2)$ as a massless,[17] spin-two particle. B_{ab} transforms under $SO(D-2)$ as an antisymmetric, second-rank tensor. The trace term Φ is a massless scalar, which is called the *dilaton* [14, p. 53].

In (2.4.2), I considered only a fixed, flat background η_{ab}. One can analyse more general possibilities by introducing the fields f_{ab}, B_{ab}, and Φ into the worldsheet action as background fields. The appropriate action is then [22, p. 429]

$$S_{\mathrm{P}(\eta+f,B,\Phi)} = -\frac{1}{4\pi\alpha'}\int_{\Sigma} d^2\sigma\sqrt{-h}\Big(h^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b\left(\eta_{ab}+f_{ab}\left(X\right)\right) + \epsilon^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b B_{ab}\left(X\right) + \alpha'\Phi R\left(h\right)\Big), \quad (2.4.3)$$

where conventions are such that $\epsilon^{\mathfrak{ab}} = \pm 1/\sqrt{-h}$. Thus, in perturbative string theory there is a move from finding the excited states of strings, to treating the associated quantum fields as *background fields* in which the dynamics of further strings can be analysed.[18]

Consider a string propagating in background fields, as per (2.4.3). This action may be regarded as that of a conformal field theory (of the X^a fields) on the string worldsheet. It turns out, remarkably, that this conformal invariance *requires for consistency* that the background fields be dynamically coupled together in the Einstein field equations, plus higher order corrections.[19] This follows since the beta function[20] associated

[17]For how the masses of such states are determined, see e.g. [14, p. 53].

[18]Attempting to make sense of this move is a philosophically nuanced task—see e.g. [99,102,175]. In this chapter, I pass over the subtleties which arise in this vicinity, and simply *assume* that it is legitimate to treat excited states of strings as background fields, as in (2.4.3).

[19]For original sources on this result, see e.g. [33]; for textbook discussion, see [90, pp. 167ff.]; for a presentation in the philosophical literature, see [102, §3].

[20]In quantum field theory, a *beta function* $\beta(g) = \partial g/\partial\log(\mu)$ encodes the dependence of a coupling parameter g upon an energy scale μ. Since a conformal field theory does not depend upon an energy scale, the beta function associated to its coupling parameters must vanish. See e.g. [159,197] for further

to each of the background fields (each such background field being treated as a coupling parameter in the worldsheet conformal field theory) can be computed perturbatively; conformal invariance then requires these beta functions to vanish. For example, for the closed bosonic string with background fields g_{ab}, B_{ab} and Φ, and $D = 26$, one finds the following one-loop beta functions associated to each of the background fields, having performed a conformal rescaling $g_{ab} \to e^{\Phi/6} g_{ab}$,[21] [33, p. 597]

$$G_{ab} = 8\pi T_{ab}, \tag{2.4.4}$$

$$\nabla_a \nabla^a \Phi + \frac{1}{6} e^{-\Phi/3} H^2 = 0, \tag{2.4.5}$$

$$e^{-\Phi/3} \nabla_a H^a{}_{bc} = 0, \tag{2.4.6}$$

with

$$T_{ab} = \frac{1}{4} \left(H_{acd} H^{cd}{}_b - \frac{1}{6} g_{ab} H_{abc} H^{abc} \right) e^{-\Phi/3} + \frac{1}{6} \left(\nabla_a \Phi \nabla_b \Phi - \frac{1}{2} g_{ab} \nabla_a \Phi \nabla^a \Phi \right), \tag{2.4.7}$$

and G_{ab} the (26-dimensional analogue of the) Einstein tensor. Analogous results for the *heterotic superstring* (in which there also exists a one-form background field A_a) yield more familiar coupled Maxwell-Einstein systems (see [33, §4]).[22]

Allowed background fields in perturbative string theory are restricted to those which correspond to the excited states of strings; one such field ($g_{ab} := \eta_{ab} + f_{ab}$,

details.

[21] Here, H_{abc} is the generalised Faraday tensor associated to B_{ab}—so that, in the notation of differential forms, $H = dB$.

[22] These results are not presented explicitly in order to avoid technical subtleties regarding superstring theories; the above will suffice for my purposes.

where f_{ab} is the graviton field) corresponds to a Lorentzian metric field; the other fields are identified as non-gravitational fields.[23] Conformal invariance of the string worldsheet metric—itself required for the consistency of the theory, and hence not an input assumption therein—guarantees that g_{ab} satisfies the Einstein field equations, and that all matter fields obey Poincaré invariant dynamical laws. Thus, in perturbative string theory, *unlike* GR, one is *not* free to choose matter field couplings; rather, the dynamical equations governing the behaviour of matter fields on target space are an ineluctable consequence of the formalism. In other words, the space of DPMs for the background fields in a given string theory has a *fixed* set of dynamical equations for matter fields. For this reason, **MR1** is, in string theory, a necessary principle. Indeed, such is also the case for **MR2**: it is again a consequence of the formalism that (in the relevant regime) the local Poincaré symmetries of the metric field g_{ab} coincide with those of the dynamical equations governing matter fields. Thus, on moving to a successor theory to GR, **MR1** and **MR2** both *cease to be miraculous*.

2.5 Logic of the Miracles

It is worth taking some time to reflect upon the logic of the status of **MR1** and **MR2** as they appear in GR, versus as they appear in perturbative string theory. In the former case, **MR1** must simply be postulated as an input assumption; **MR2** then follows once it is noted that there also exists in GR an autonomous Lorentzian metric field, the symmetries of which happen to coincide with the symmetries of the dynamical laws governing matter fields in the theory, assuming **MR1**. Thus, the situation in GR is as follows:

[23] Recall from Ch. 1 that I regard 'matter'/'non-gravitational' fields as being those with an associated stress-energy tensor, and 'gravitational' fields as being those with no associated stress-energy tensor.

$$\begin{aligned} & \textbf{MR1} \\ \textbf{MR1} + \text{Lorentzian } g_{ab} \text{ in KPMs} \quad \Rightarrow \quad & \textbf{MR2} \end{aligned}$$

Contrast this with the case of perturbative string theory. In that case, one posits a Minkowski metric field η_{ab} on the target space, but having done so **MR1** and **MR2** follow as consequences of the formalism, with no additional input assumptions. Thus, the situation in perturbative string theory is as follows:

$$\eta_{ab} \text{ in KPMs} \quad \Rightarrow \quad \textbf{MR1} + \textbf{MR2}$$

Clearly, this is not to say that there are *no* input assumptions in perturbative string theory at all; it is, rather, to say two things: (a) the *number* of input assumptions in perturbative string theory versus GR is reduced; (b) what were two unexplained coincidences in GR (*viz*., **MR1** and **MR2**) admit of a deeper, overarching explanation in the case of perturbative string theory.[24]

To put matters differently, in the case of GR, **MR2** consists of two points: (i) a restatement of **MR1**, and (ii) the observation that one is dealing with a Lorentzian metric field g_{ab} in the KPMs of the theory. Clearly, it is (ii) that is the extra component of **MR2**, over and above **MR1**. But note that this extra assumption is analogous to the sole input assumption in string theory, that one is dealing with a Minkowski metric field η_{ab} in the KPMs of the theory.[25] In the case of string theory, this extra assumption

[24]Setting up the logic of the two miracles in this way raises an interesting question: would it be the case that one would recover the analogues of **MR1** and **MR2** in string theories with the structure of *Galilean* spacetime on the target space (cf. e.g. [88])? Absent the calculations, it is hard to be concrete on this point—though I agree that this is an interesting question worthy of future pursuit.

[25]In fact, this latter assumption is somewhat stronger, for in the case of string theory one assumes

suffices, ultimately, to account for **MR1**, whereas such is *not* the case in GR. Thus, while in GR there exist two input assumptions, in string theory there exists only one.

2.6 Emergent Spacetime

The observation that **MR1** and **MR2** *must* hold in perturbative string theory has important implications for the emergence of spacetime in quantum gravity. Drawing upon Knox's discussion at [111, §1], distinguish two senses in which the 'emergence of spacetime' might be understood. On the former, such emergence arises whenever an object (or collection of objects) in a given theory is not fundamental—in the sense of its not being defined at the level of KPMs of the theory, but rather being constituted out of those fundamental objects specified in the KPMs, in a possibly limited range of physical circumstances—but which nevertheless plays the functional role of spacetime as characterised in §2.3. Knox claims (correctly) that one need not look to theories of quantum gravity to find illustrations of this phenomenon—for example, teleparallel gravity (an alternative theory of classical gravity to GR, in which curvature degrees of freedom are put into torsion in the connection—for Knox's discussion of this theory, see [110, 111]; for further philosophical details, see [175, 183]) seems to instantiate just such a point, for in this theory those spatiotemporal degrees of freedom are composite objects, built from fundamental fields specified at the level of KPMs.[26]

The second sense in which spacetime can be emergent is the following: a given object's playing the functional role of spacetime *à la* Knox is not a necessary conse-

a specific form of the target space metric field (namely, that it is the Minkowski metric), whereas in GR one assumes only that this metric field be Lorentzian. Many thanks to Nick Huggett for useful discussion on this point.

[26] Another classical theory which exemplifies the emergence of spacetime in this sense is Bekenstein's bimetric T*e*V*e*S ('Tensor-Vector-Scalar') theory, presented in Ch. 1.

quence of the formalism, in the sense that this object (fundamental or not) plays that functional role in all models of the theory, but is rather contingent, in the sense that such an object plays the spacetime role in some models of the theory, but not others. Again, Knox correctly observes at [111, §1] that one does not need to look to theories of quantum gravity for illustrations of this case—GR, for the reasons outlined in §2.3, constitutes a clear example of this latter sense of emergence.

Though GR exemplifies the emergence of spacetime in the latter sense, it does *not* exemplify it in the former sense, for in this theory it is the *fundamental* metric field g_{ab} which plays the functional role of spacetime. The situation is very different in perturbative string theory, in which the object playing the functional role of spacetime is the composite g_{ab}, built out of the Minkowski metric field η_{ab} and graviton field f_{ab}, the latter itself a coherent state of strings in an excited graviton mode.[27] Since this object is non-fundamental, it qualifies as emergent spacetime in the former sense; however, it does *not* qualify as emergent spacetime in the latter sense, for this object *necessarily* plays the functional role of spacetime, in all models of the theory. In other words, GR and perturbative string theory are, in a certain sense, 'dual' *vis-à-vis* emergent spacetime in these theories, for while spacetime in GR is *fundamental but contingent*, spacetime in perturbative string theory is *non-fundamental but necessary*.

Now, it is worth noting that perturbative string theory does not, of course, exhaust the string-theoretic picture. While the above story might be correct for perturbative string theory, it might *not* hold in general for a *non-perturbative* formulation of the theory. For example, it has been conjectured that some solutions of 'M-theory'—perhaps the best-known candidate for a non-perturbative formulation of string theory—will not rely, fundamentally or otherwise, upon spatiotemporal con-

[27] For some discussion of the extent to which the η_{ab} field can be eliminated in perturbative string theory, see [175, ch. 8].

cepts.[28] If that is correct, then it may be that the idea of spacetime being necessary in non-perturbative string theory does not hold water, unlike the situation in the perturbative case. Though these observations do not undermine the claims of the current chapter, they certainly do deserve to be flagged.

2.7 Dualities

Up to this point, I have taken it for granted that, on a spacetime functionalist approach, the target space metric field $g_{ab} := \eta_{ab} + f_{ab}$ represents spacetime. However, at this juncture I must acknowledge recent work by Huggett, in which it is argued that this target space metric field cannot represent "phenomenal spacetime", i.e., "the geometrical space we take ourselves to experience in the everyday, including the experience of three large dimensions" [100, p. 7]. The reason for this has to do with the phenomenon of *T-duality* in string theory.[29]

T-duality is just one of a broad class of 'dualities' in theoretical physics in general, and string theory in particular. For the purposes of this thesis, a *duality* can be taken to be a mapping between the spaces of DPMs of two theories, such that any two models related by that mapping are empirically equivalent (in turn, the empirical equivalence of two models can be defined in terms of their agreeing on all *empirical substructures*, in the sense of van Fraassen [210, pp. 67ff.]).[30] In the case of T-duality, models of perturbative string theory on a target space product manifold $M \times S^1$ with radius of the periodic dimension R are found to be dual to models of perturbative

[28]The question of whether M-theory may or may not deploy spatiotemporal concepts was the sixth open question on string theory presented by Strominger at the *Strings 2014* meeting—cf. [201]. For some philosophical discussion of this matter, see [133, p. 14].

[29]See [103, §2.4] for related discussion, and [133] for an extension of Huggett's arguments to other string-theoretic dualities.

[30]For more on the definition of dualities, see e.g. [48, 51, 184]; dualities will constitute the focus of Part III of this thesis.

string theory on the target space product manifold $M \times S^1$ with radius of the periodic dimension equal to α'/R (see e.g. [14, ch. 6]).

Within the literature on dualities, some (e.g. [184, 187]) have taken it that duality-related models should be regarded *ab initio* as representing the same possible world (i.e. as being not just empirically equivalent, but also *physically* equivalent); others (e.g. [48]) have resisted this conclusion. (I fall into the latter camp—see Ch. 6.) Assuming that one endorses the former approach, then, since duality-related models often have very different mathematical structures, there exists a burden to account for the common ontology which such models are taken to represent; one popular option is to seek an explication in terms of the mathematical structure *common* to both models (cf. Ch. 5). Not only does Huggett fall into the former interpretative camp—that is, he takes it that duality-related models should be regarded *ab initio* as representing the same physical state of affairs[31]—but, in addition, he maintains that the physical state of affairs represented by such duality-related models should be explicated in terms of their common structure [100, §2.3].

In light of this, Huggett's position on the interpretation of T-dual models of perturbative string theory can be summarised as follows. While "[i]t is natural when first introduced to string theory, to think that target space is simply the same space we ordinarily experience, or at least space as conceived in contemporary physics" [100, p. 8], ultimately "T-duality shows that a definite radius for target space ... [is] not physical, but only [an artefact] of the representation" [100, p. 10], for this radius is not part of the structure common to the two models under consideration. Since this aspect of the target space metric field is unphysical, for Huggett, it follows that the target space metric field g_{ab} cannot always represent phenomenal spacetime; rather, what *does* represent phenomenal spacetime must be constructed from the shared structure

[31] At least for theories which are supposed to constitute descriptions of the entire universe—see [100, §2.2], and [48, §3.1] for further discussion.

of the T-dual models. In fact, following arguments by Brandenberger and Vafa [23], Huggett maintains that phenomenal spacetime always coincides with the target space of the *larger* radius between two T-dual models.[32,33]

If Huggett is correct here, then does this pose a problem for our verdict that spacetime is 'non-fundamental but necessary' in perturbative string theory? On the one hand, one might be inclined to answer this question in the negative, for (one might say) there is nothing to stop an object in a theory playing the functional role of spacetime, while at the same time not being a candidate for representing phenomenal spacetime, or ultimately being regarded as 'surplus structure'. Nevertheless, one might feel that there is more to be said here, and that an affirmative answer to the above question is also possible. In particular, Huggett's discussion of phenomenal space as being that construct which is surveyed by *macroscopic* rods and clocks might make one think that whatever plays the functional role of spacetime should *also* satisfy this criterion. In that case, one should be disturbed by the verdict of Knox's spacetime functionalism, that what plays the functional role of spacetime need not coincide with phenomenal spacetime in this sense.

If one buys this reasoning, then one might read this discussion of T-duality as highlighting that Knox's functional definition of spacetime as that structure which "defines a structure of local inertial frames" cannot be *sufficient* for the identification

[32]Note, though, that Huggett does not maintain that phenomenal spacetime *just is* the target space of the larger radius, for the model under consideration with the *smaller* radius will also (of necessity, in light of the duality) manifest the mathematical structure needed to represent phenomenal spacetime, while not exhibiting fundamentally a target space of large radius. Thanks to Nick Huggett for stressing this point.

[33]There is some tension between this claim that, in the case of T-dual models, phenomenal spacetime always coincides with the target space of the larger radius, and Huggett's earlier claim that phenomenal spacetime is "the geometrical space we take ourselves to experience in the everyday, including the experience of three large dimensions" [100, p. 7]—for the former, at least as presented above, will have more than three "large" spacetime dimensions. Here is one way to resolve the tension: by "phenomenal spacetime", Huggett means the classical structures which represent the spatiotemporal data which we *would* experience, *were we* to be 'embedded' in the model under consideration. If that is right, then (e.g.) having three dimensions is not, in fact, necessary for a given structure to qualify as "phenomenal spacetime".

of spatiotemporal structure, for if what qualifies functionally as spacetime must *also* qualify as phenomenal spacetime, then this object should, as Huggett notes, also be surveyed by macroscopic rods and clocks. Only one of the T-dual target space metric fields does as much; thus, with this extra criterion imposed, our above conclusions must be revised, and we should—it appears—conclude that spacetime in perturbative string theory is neither fundamental nor necessary, in the senses of §2.6.

In fact, however, the situation on these matters is even more subtle, for one can argue that, in any model of T-dual perturbative string theories, one can (in light of the formal duality map) always identify *some* structure corresponding to phenomenal spacetime. It that is correct, then one may be able to argue that every model of T-dual string theory *does* possess some structure corresponding to spacetime, even on the above-proposed revised Knoxian account. In that case, spacetime in T-dual perturbative string theories would be non-fundamental but necessary after all. Indeed, it would be *doubly* non-fundamental, for not only are the target space metric fields g_{ab} composite objects, but, in addition, what represents (phenomenal) spacetime should not (as Huggett stresses) be identified straightforwardly with those objects.

Let me recapitulate the dialectic of this section. First, I observed that Knox's original functionalist approach to spacetime would, it appears, identify the target space metric fields g_{ab} in models of perturbative string theory as being spatiotemporal. On that view, spacetime in perturbative string theory is non-fundamental, but necessary. I then observed (following Huggett), however, that target space metric fields need not be identified as spatiotemporal, for they need not be surveyed by macroscopic rods and clocks. If one augments Knox's functional definition of spacetime to include also this criterion, then (it appears—so long as one continues to focus one's attention upon the target space metric fields) that spacetime in perturbative string theory is non-fundamental and contingent. Finally, however, I reflected further upon phenomenal spacetime. Since every model of T-dual perturbative string theory possesses sufficient mathematical structure to represent phenomenal spacetime (in light of the existence of the duality map—so even those models of perturbative string theory with *small* target space radius possess *some* structure corresponding, via the duality map, to a target space of *large* radius, which coincides with phenomenal spacetime), and since this structure does qualify as spatiotemporal on our proposed augmented Knoxian functional definition of spacetime, it seems that, ultimately, even on this revised functional definition, spacetime in perturbative string theory is non-fundamental but necessary after all.

2.8 Conclusions

A number of tasks have been accomplished in this chapter. Having recalled from Ch. 1 **MR1** and **MR2** and their 'miraculous' status in GR in §2.2, and their connections with the chronogeometricity and spatiotemporality of the metric field in §2.3, I turned to perturbative string theory in §2.4. In that section, I demonstrated that **MR1** and **MR2** cease to be miraculous in this theory of quantum gravity—rather, they are necessary consequences of the formalism. This done, I considered the chronogeometricity and spatiotemporality of the target space metric field g_{ab} in perturbative string theory; since **MR1** and **MR2** hold of necessity in this theory, so too (on a Knoxian account) does the spatiotemporal status of g_{ab}. I clarified the logic of these arguments in §2.5.

In §2.6, I distinguished two senses of 'emergent spacetime': (A) the object which plays the spatiotemporal role in a given theory not being part of the fundamental ontology of that theory (as specified in its KPMs); versus (B) the object which plays the spatiotemporal role doing so in some, but not all, models of the theory. I argued that while (again on a Knoxian account) GR exemplifies (B) but not (A), perturbative string theory exemplifies (A) but not (B). Finally, in §2.7 I drew upon Huggett's discussion of T-duality and phenomenal space to reflect upon whether these verdicts are reasonable, and whether both they, and Knox's spacetime functionalism more generally, should be revised.

There are several important upshots of this work. First, (i) the coincidence of inertial and gravitational mass, or (ii) the geodesic principle, are arguably 'miraculous'—unexplained—in theories prior to GR, but find natural explanation in that context.[34] The observation that **MR1** and **MR2** admit of natural explanation in perturbative string theory, where none was forthcoming in GR, can be seen as a continuation of this trend of accounting for apparent coincidences in one's physics by recourse to a more 'fundamental' theory. Second, this work is useful for clarifying senses of emergent spacetime, some of which apply to GR, and others to string theory; this illustrates that there is no univocal notion of the 'emergence of spacetime' in a given theory. Third, consideration of phenomenal spacetime in Huggett's sense re-emphasises that more needs to be done to understand fully the spatiotemporal picture with which perturbative string theory presents us.

Possible extensions of investigations into the extent to which **MR1** and **MR2** admit of explanation in successor theories to GR abound. Essentially, what one needs in order to account for these 'miracles' is the existence of a unified origin for dynamical and metric symmetries. There are some hints that this can be attained not just in perturbative string theory, but also via 'supergravity' and 'double field theory'.[35] The exploration of such matters should constitute a philosophically fertile research programme for the future.

[34] See e.g. [24, 219] for discussion of such matters. For some discussion calling into question such claims regarding the geodesic principle in particular, see [127, 227, 228].

[35] Many thanks, respectively, to Richard Dawid and Nic Teh for these suggestions. For introductions to these areas of theoretical physics, see respectively [148] and [2].

Chapter 3

Explanation, Geometry, and Conspiracy in Relativity Theory

I discuss the debate between dynamical versus geometrical approaches to spacetime theories, in the context of both special and general relativity. In Ch. 1, I argued that this debate takes a substantially different form in special versus general relativity, and that there exist problems for a certain, strong version of the geometrical approach. I now extend this discussion in several respects. First, I argue that different versions of the geometrical approach—only some of which are viable—should be distinguished. Second, I argue that, in general relativity, there is no difference between the most viable version of the geometrical approach (which I call the 'qualified' geometrical approach) and the dynamical approach. Third, I demonstrate that the two 'miracles' of general relativity admit of no resolution from within general relativity, on either the dynamical or 'qualified' geometrical approaches, modulo some possible hints that the second 'miracle' may be resolved by appeal to recent results regarding the 'geodesic principle' in general relativity.

3.1 Introduction

It is roughly a decade since the groundbreaking work of Brown and Pooley [24,27,28] brought into the mainstream philosophy of physics literature the debate between *dynamical* versus *geometrical* approaches to spacetime theories—a debate which has already been introduced in Ch. 1, but which I will now address more directly. At the most general level, this debate centres upon the following question: *whence the chronogeometric significance of the metric field?* That is, why is the metric field (in theories such as special and general relativity) surveyed by rods and clocks built from matter fields? While the geometrical approach maintains that the metric field (in some sense) *explains* or *constrains* the form of the dynamical laws for matter fields, such that those fields behave such as to survey the metric field, advocates of the dynamical approach, by contrast, claim that an account of the chronogeometric significance of the metric field may begin from considerations regarding only the dynamical laws governing matter fields themselves.

Of course, this is vague; in §3.3 of this chapter, I sharpen significantly the above presentation of the debate. Nevertheless, even at this early stage, a number of genuine and substantial questions arise:

1. Does the dynamical/geometrical debate take the same form in the context of theories with fixed metric structure (such as special relativity (SR)) as it does in theories with dynamical metric structure (such as general relativity (GR))? (In Ch. 1, we have seen that the answer to this question is negative; in the current chapter, I will expand this account.)

2. What notion of explanation is at play in this debate? Does answering this question reveal multiple different senses in which the dynamical/geometrical ap-

proaches may be understood?

3. Are the dynamical and geometrical approaches truly distinct from one another at all?

4. How does the dynamical approach relate to e.g. the 'spacetime functionalism' of Knox [111–113], or recent discussion on these matters by Weatherall? [228, §6]

In brief, my answers will run as follows. On (1), I argue in §3.3 of this chapter—developing upon Ch. 1—that there exist *significant* differences regarding this debate as it occurs in the context of SR, versus as it occurs in GR. The principal reason for this is that, while the advocate of the dynamical approach may be regarded as seeking to *ontologically reduce* the metric field in theories with fixed metric structure (such as SR) to the symmetry properties of matter fields (cf. [29, 147]), she does *not* attempt to make such a move in theories with dynamical metric structure, such as GR.[1]

On (2), I argue in §3.4 that it is important to distinguish between what I call 'qualified' versus 'unqualified' explanations in the context of this debate.[2] Once this distinction is made, the geometrical approach bifurcates into two positions, which I call, respectively, the 'qualified' and 'unqualified' geometrical approaches. In §3.5, I argue that distinguishing between these two positions is crucial, for while the former version of the geometrical approach is tenable, the latter is *not*.

[1] In light of the fact that the advocate of the dynamical approach does not attempt to undertake an ontological reduction of the above-described kind in the context of GR, one might be inclined to conclude: 'So much the worse for the dynamical approach in the context of GR, as a distinct view in the landscape'. Below, I will argue that there is something to this concern, for (I maintain) there is no difference in the GR context between the dynamical approach and the most defensible version of the geometrical approach.

[2] It is worth flagging that I will offer these two notions of explanation without claiming (or seeking) to give a full conceptual analysis of the notion of scientific explanation; in my view, the distinction between 'qualified' and 'unqualified' explanations is still a valuable and comprehensible one (providing, as I see it, at least some of the "explanatory concepts" which Norton suggests may be necessary for "a full understanding of constructivism [i.e., the dynamical approach]" [153, p. 824]), even in the absence of such an analysis. (In this regard, cf. the methodology of [227, pp. 15-16].)

On (3), I argue in §3.5 that, while the 'unqualified' version of the geometrical approach is distinct from the dynamical approach in the context both of theories such as SR and of theories such as GR, the 'qualified' geometrical approach, by contrast, is only distinct from the dynamical approach in the former context.

On (4), I argue that there is an important sense in which Knox's spacetime functionalism, according to which "the spacetime role is played by whatever defines a structure of local inertial frames" [113, p. 22], constitutes an *extension* of the dynamical approach—in essence stating that whichever structure has chronogeometric significance may be identified as playing the functional role of spacetime, and therefore, on a functionalist approach to the definition of physical quantities, may be identified as being spatiotemporal *tout court*.[3] In addition, I argue that Weatherall in [228, §6] is most plausibly read as both (a) embracing spacetime functionalism, and (b) embracing either the dynamical or the 'qualified' geometrical approach.

Along the way, a number of other tasks are accomplished. Most notably, I demonstrate that what were labelled in Ch. 1 the two 'miracles' of GR—(1) that all dynamical laws for matter fields have the same local (Poincaré) symmetry properties; and (2) that these local symmetries coincide (in the relevant regime in which curvature effects may be ignored) with the symmetries of the ontologically autonomous metric field in the theory—admit of no resolution from within GR, on any plausible form of the dynamical or geometrical approaches, modulo some hints from recent work on the so-called 'geodesic principle' in GR regarding the second 'miracle'.

[3] Cf. Ch. 2. By contrast, there is a sense in which advocates of the dynamical approach need not speak of 'spacetime' at all—cf. [30, §3.1].

3.2 Background

Before proceeding to the matters outlined above, I review in this section some standard discussion regarding (i) the formulation of classical spacetime theories (§3.2.1); (ii) symmetries in such theories (§3.2.2); and (iii) presentations of special and general relativity (§3.2.3). Though there is some repetition here from the preceding chapters, the level of detail of the presentation in the current chapter is higher—which will be important in the ensuing.

3.2.1 Spacetime theories

Let us say—following e.g. [5,169,171,207]—that the *kinematically possible models* (KPMs) of a given spacetime theory are picked out by tuples $\langle M, \Phi_1, \ldots, \Phi_n \rangle$, with (a) M a (four-dimensional) differentiable manifold; and (b) the $\Phi_1, \ldots, \Phi_n$ various (tensor) fields on M.[4] Given a class of KPMs for a given theory, let us then say that the *dynamically possible models* (DPMs) of that theory are those KPMs the $\Phi_1, \ldots, \Phi_n$ of which satisfy certain specified dynamical equations.

To illustrate, consider two examples. First, take a special relativistic massless Klein-Gordon theory (call it **KGS**). In this theory, KPMs are triples $\langle M, \eta_{ab}, \varphi \rangle$, where η_{ab} is a fixed Minkowski metric field on M (fixed *identically* in all KPMs—see [171, p. 115]), and φ is a real scalar field on M. DPMs of **KGS** are picked out as those

[4] In principle, we should not exclude other types of field on M—e.g. spinor fields; pseudotensors; tensor densities; etc. (For arguments for taking these latter two classes of object seriously, see [160, 161].) In this chapter, however, I focus exclusively upon the case in which the Φ_i are tensor fields.

KPMs the fields of which satisfy the massless Klein-Gordon equation,[5]

$$\eta_{ab}\nabla^a\nabla^b\varphi = 0. \tag{3.2.1}$$

As a second example, consider a *general* relativistic Klein-Gordon theory (call it **KGG**). In this case, KPMs are again triples $\langle M, g_{ab}, \varphi\rangle$—this time, however, g_{ab} is a generic Lorentizan metric field on M, *not* fixed in all DPMs of the theory. In this case, DPMs are picked out by the GR Klein-Gordon equation,[6]

$$g_{ab}\nabla^a\nabla^b\varphi = 0, \tag{3.2.2}$$

and the Einstein field equations,

$$G_{ab} = 8\pi T_{ab}, \tag{3.2.3}$$

where T_{ab} is the stress-energy tensor associated with φ.

3.2.2 Symmetries

I now draw a standard distinction between *metric symmetries*, and *dynamical symmetries* (cf. e.g. [64, §3.4]).

For a given metric field, let us say that a coordinate transformation is a *metric symmetry* (sometimes: an *isometry*) just in case the metric field is unaltered by the co-

[5]Here, ∇_a is the torsion-free derivative compatible with η_{ab}, so that $\nabla_a\eta_{bc} = 0$.
[6]The torsion-free derivative operator ∇_a now compatible with g_{ab}, so that $\nabla_a g_{bc} = 0$.

ordinate transformation. For example, the symmetries of the Minkowski metric field η_{ab} of special relativity are the *Poincaré transformations*—those affine transformations[7]

$$x^\mu \to \Lambda^\mu{}_{\mu'} x^{\mu'} + a^\mu \tag{3.2.4}$$

the linear transformation matrix components $\Lambda^\mu{}_{\mu'}$ of which satisfy

$$\Lambda_{\mu'}{}^\mu \Lambda_{\nu'}{}^\nu \eta_{\mu\nu} = \eta_{\mu'\nu'}, \tag{3.2.5}$$

and which are hence *Lorentz transformations*. By contrast, the metric field g_{ab} of GR need not in general have any non-trivial symmetries—although it might, in particular models of the theory (cf. Ch. 4).

In addition to the notion of metric symmetries, it is useful to introduce the notion of a *dynamical symmetry*. A coordinate transformation is a *dynamical symmetry* just in case the dynamical equations governing non-gravitational fields take the same form in coordinate systems related by that transformation. For example, transforming the SR Klein-Gordon equation (3.2.1) from one coordinate system to a second via an affine transformation[8]

$$\eta_{\mu\nu} \nabla^\mu \nabla^\nu \varphi = 0 \tag{3.2.6}$$

$$\longrightarrow \quad \eta_{\mu\nu} M^\mu{}_{\mu'} M^\nu{}_{\nu'} \nabla^{\mu'} \nabla^{\nu'} \varphi = 0, \tag{3.2.7}$$

[7]Here, I switch to a coordinate-based description—hence the transition from Roman (abstract) to Greek indices.

[8]Here, again, I use a coordinate-based description; in this case the linear transformation matrix components in this affine transformation are written $M^\mu{}_{\mu'}$, for reasons which shall become clear. Note that I do not transform the fixed fields—cf. [171, p. 115].

one finds that such an equation is invariant under the transformation when (3.2.5) is satisfied—i.e. if the affine transformation is a Poincaré transformation. Thus, the dynamical symmetries associated with (3.2.1) at least include the Poincaré transformations.

3.2.3 Special and general relativity

Having introduced the necessary details regarding spacetime theories and their symmetries, in this section I characterise—with both greater precision and generality—what it means for a given theory to be 'special relativistic' (§3.2.3.1), versus 'general relativistic' (§3.2.3.2).

3.2.3.1 Special relativity

In this chapter, I take special relativistic theories to be characterised by the following two criteria:

- KPMs $\langle M, \eta_{ab}, \Phi_1, \ldots, \Phi_n \rangle$ at least include a fixed Minkowski metric field η_{ab}.
- DPMs are picked out by the requirement that dynamical equations for the $\Phi_1, \ldots, \Phi_n$ be Poincaré invariant.

The latter criterion is referred to by Brown as the *big principle*—see e.g. [24, §8.4.1].[9] Note that, by construction, metric and dynamical symmetries coincide in special relativistic theories.

[9]In his 1908 paper [139], Minkowski referred to this principle as the *world-postulate*—for discussion, see [24, §8.1].

3.2.3.2 General relativity

Turn now to the question of what it is for a spacetime theory to be general relativistic. For the purposes of this chapter, I take such theories to be characterised by the following two criteria:

- KPMs $\langle M, g_{ab}, \Phi_1, \ldots, \Phi_n \rangle$ at least include a Lorentzian metric field g_{ab}.
- DPMs are picked out by dynamical equations for the $\Phi_1, \ldots, \Phi_n$, along with the Einstein field equations $G_{ab} = 8\pi T_{ab}$, where T_{ab} is the stress-energy tensor for the $\Phi_1, \ldots, \Phi_n$.

This characterisation of a general relativistic theory is very weak—note, in particular, that there is no guarantee in general relativistic theories so understood that metric symmetries coincide locally with dynamical symmetries (as was the case for special relativistic theories, as presented above).[10] Such requirements may be imposed via restriction to those models of GR which satisfy further conditions; from the point of view of the matter of symmetry coincidence, one particular auxiliary condition which will be of significance is:[11]

- Instantiation of the strong equivalence principle (SEP).

Whence this third assumption? What exactly is the SEP, and why need it be imposed in one's characterisation of a general relativistic theory? A full answer to these

[10]One further observation about the distinction between special versus general relativistic theories as characterised above: Since the metric field η_{ab} of SR is fixed identically in all KPMs, so too is the manifold M on which that field is defined. Not so for GR: since it is not definitional of a general relativistic theory that it contain a certain fixed field, there may exist models with distinct manifolds M.

[11]Other conditions which one may be interested in imposing upon the class of GR solutions in which one is interested are e.g. *energy conditions*, for such conditions are often understood to be tied to the restriction to 'physically reasonable' matter (for example, to conditions that energy cannot be negative). For a recent virtuoso study of energy conditions, see [41].

questions will require some detailed discussion. (For more on the SEP, cf. Ch. 1.)

The SEP is intended to capture facts regarding the 'local validity' of SR in GR. Recall from Ch. 1 that Brown puts the point thus:

> There exists in a neighbourhood of each event preferred coordinates, called *locally inertial* at that event. For each fundamental non-gravitational interaction, to the extent that tidal gravitational forces can be ignored, the laws governing the interaction find their simplest form in these coordinates. This is their *special relativistic form*, independent of spacetime location. [24, p. 169]

Here, there exist a number of subtleties regarding what is meant by the qualification "to the extent that tidal gravitational forces can be ignored", and moreover regarding whether other foundational principles GR—for example *minimal coupling*, which is a heuristic prescription for the construction of dynamical laws for non-gravitational fields in GR from those in SR—are compatible with the SEP as formulated above. These issues have already been discussed in Ch. 1. In this chapter, the essential aspect of the SEP is the imposition that, in the neighbourhood of any $p \in M$ in GR, laws of physics recover their 'special relativistic form'—where I shall understand this to mean: *a Poincaré invariant form*. Clearly, this is a particular restriction on the matter sector in the theory.

To illustrate, consider again the general relativistic Klein-Gordon equation (3.2.2). Written in an arbitrary coordinate basis, this reads

$$g_{\mu\nu}\partial^{\mu}\partial^{\nu}\varphi + \Gamma^{\mu}{}_{\nu\mu}\partial^{\nu}\varphi = 0. \tag{3.2.8}$$

In normal coordinates at $p \in M$ (the details of which were presented in Ch. 1), this takes a particularly simple form at p:

$$\eta_{\mu\nu}\partial^{\mu}\partial^{\nu}\varphi = 0; \tag{3.2.9}$$

moreover, this form (with the metric field diagonalised) is retained in all frames related by Poincaré transformations. This illustrates the sense in which certain dynamical equations for non-gravitational fields recover locally a Poincaré invariant form. That *all* dynamical laws for non-gravitational fields in GR manifest this quality is a statement of the SEP. Importantly, note that, absent the imposition of the SEP, it is *not* the case that all dynamical equations for matter fields in GR need be locally Poincaré invariant. For example, there exists no *a priori* prohibition on the existence of matter fields obeying dynamical laws which are locally *Galilean* invariant, in a spacetime theory with a dynamical, Lorentzian metric field satisfying the Einstein field equations.[12]

Why should one restrict to those solutions of GR in which the SEP is satisfied? The reason is that this principle—which ensures that, locally, the (Poincaré) symmetries of the metric field[13] coincide with the (Poincaré) symmetries of the dynamical laws governing matter fields—is typically regarded to constitute an important condition for the *chronogeometricity* of the metric field—that is, for intervals as given by the metric field to be read off by stable rods and clocks built from matter fields. That is, the SEP is, it is argued, an important condition for the metric field g_{ab} in GR to have (local) *operational meaning*. These matters have already been discussed in detail in Ch. 1. I return to discuss further the SEP, in light of the so-called *geodesic principle*

[12] This possibility was already raised in Ch. 2.

[13] Again, modulo subtle issues regarding the qualification "to the extent that tidal gravitational forces can be ignored"—see Ch. 1 for discussion.

in GR, in §3.6.

3.3 The dynamical/geometrical debate

The above in hand, in this section I demonstrate how the dynamical/geometrical debate plays out in the context of both SR (§3.3.1) and GR (§3.3.2);[14] a detailed reconsideration of the geometrical approach will follow in §§3.4-3.5. At the most general level, the dynamical/geometrical debate centres upon the following question:

> Whence the metric field's chronogeometric significance?

Taking, as elaborated in Ch. 1, the (local) coincidence of metric and dynamical symmetries to be an important condition which must be fulfilled in one important means via which the metric field acquires its chronogeometric significance (*viz.*, via the SEP), one antecedent question which one might seek to address in order to answer the above is the following:[15]

> Why do metric symmetries coincide (locally) with dynamical symmetries?[16]

It is upon this latter question that much of the dynamical/geometrical debate has focussed. *Prima facie*, the *dynamical approach* (developed in particular by Brown and Pooley [24,27,28]) appears to offer a very different account of this coincidence of symmetries to the *geometrical approach* (advocated by, for example, Friedman and Maudlin

[14]There is a sense in which the lessons of §§3.3.1 and 3.3.2 can be generalised to *all* theories with, respectively, fixed versus dynamical metric structure—see [30, §5].

[15]For the time being, my focus is on this mode of gaining operational access to the metric field—though I concede that there may be other means, as discussed in §3.6 below.

[16]The 'locally' qualification is of particular significance in GR, since the SEP ensures the *local* coincidence of metric and dynamical symmetries, in the neighbourhood of a given point $p \in M$.

[81, 137]). In the remainder of this section, I discuss the dynamical/geometrical debate in the context of both SR (§3.3.1) and GR (§3.3.2).[17]

3.3.1 Special relativity

In this subsection, I consider the account of the coincidence of metric and dynamical symmetries in SR proffered on the part of advocates of the geometrical (§3.3.1.1) and dynamical (§3.3.1.2) approaches.

3.3.1.1 The geometrical approach

Why, in special relativistic theories, do dynamical symmetries coincide with symmetries of the Minkowski metric field η_{ab}? Advocates of the *geometrical approach* to spacetime theories seek to answer this question via some appeal to η_{ab} itself. To be specific, in this chapter I focus upon a version (later: versions) of the approach according to which the Minkowski metric field η_{ab} of SR is *ontologically autonomous and primitive*, and (somehow; in some sense to be cashed out) *constrains* the possible form of dynamical equations for matter, such that metric symmetries coincide with dynamical symmetries. Recall again the following passage from Maudlin: [137, pp. 117-8]

> ... the Minkowski geometry takes exactly the same form described in [any] Lorentz coordinate system (by the symmetry of Minkowski spacetime), and the laws of physics take exactly the same coordinate-based form when stated in a coordinate-based language in any Lorentz coordinate system (*because the laws can only advert to the Minkowski geometry*, and it has the

[17] In Ch. 1, the geometrical and dynamical approaches were sometimes referred to, respectively, as the '(A)-view' and '(B)-view', respectively.

same coordinate-based description). (My emphasis.)

That a notion of constraint is at play on this view is manifest in the italicised portion of the above quotation. While advocates of the dynamical approach often object that such a notion of constraint or explanation is mysterious—for example, Brown writes "It is wholly unclear how this geometrical explanation is supposed to work." [24, p. 134]—I will assess in §3.4 the extent to which such objections find their mark against all possible versions of the geometrical approach. In the meantime, I turn to the dynamical approach to SR.

3.3.1.2 The dynamical approach

The dynamical approach offers a very different perspective on the coincidence of metric and dynamical symmetries in SR. According to this view, the metric field η_{ab} is not ontologically autonomous and primitive; rather, it is a *codification* of the symmetry properties of the dynamical equations governing matter fields. (One may, therefore, understand the dynamical approach to SR—and to theories with fixed metric structure more generally—as an ontological thesis; as a form of *relationalism*—cf. [169, §6.3.2].) As Brown puts it: (Cf. also [28, p. 80], and Ch. 1, where this quote was also given.)

> The appropriate structure is Minkowski geometry *precisely because* the laws of physics of the non-gravitational interactions are Lorentz covariant. [24, p. 133]

In other words (those of Myrvold), on the dynamical view,

> [T]he connection between spacetime [metric] structure and dynamical symmetries and asymmetries is analytic. [147, p. 13]

If such a view regarding the analytic connection between metric and dynamical symmetries can be made to hold together, then that metric and dynamical symmetries coincide in SR follows *automatically*; in this way, a straightforward account of this coincidence is, apparently, available.

The question of whether the dynamical approach to SR is viable has been widely discussed—see e.g. [1,24,98,108,153,169,199,200]. In this chapter, I focus on a different issue: whether 'geometricians' can, in fact, offer a coherent answer to the question of why metric and dynamical symmetries coincide, in SR. Before doing so, however, I consider how the nature of the dynamical/geometrical debate shifts on moving to GR.

3.3.2 General relativity

Recall from Ch. 1 that in the GR context, advocates of both the dynamical and geometrical approaches *agree* that the metric field g_{ab} is an ontologically autonomous entity, obeying its own dynamical equations, and not straightforwardly reducible to (symmetries of dynamical equations governing) matter fields, as per the dynamical approach to SR.[18] However, the two approaches *prima facie* continue to issue different verdicts on the question of why metric and dynamical symmetries may be taken (locally) to coincide. In this subsection, I review the geometrical (§3.3.2.1) and dynamical (§3.3.2.2) approaches to GR.

[18] This said, the question of whether an ontological excision of the metric field in GR is possible remains of philosophical and conceptual interest—particularly to advocates of the dynamical approach, for whom this would afford a means of bringing their approach to GR into line with their approach to SR.

3.3.2.1 The geometrical approach

Advocates of the version (later: versions) of the geometrical approach to GR under consideration in this chapter maintain that, locally in the neighbourhood of any $p \in M$ (and in the regime in which 'tidal gravitational forces' may be ignored—cf. §3.2.3.2), metric and dynamical symmetries coincide (in accordance with the SEP), because the metric field g_{ab} (somehow; in some sense to be cashed out) *constrains* the possible form of dynamical equations for matter, such that metric symmetries coincide (locally) with dynamical symmetries. While, again, the advocate of the dynamical approach may find the notion of constraint here mysterious, I discuss in §§3.4-3.5 the extent to which these matters can be accounted for by advocates of the geometrical approach.

3.3.2.2 The dynamical approach

Assuming that the metric field in GR is not ontologically reducible to (symmetries of dynamical laws governing) matter fields, the foregoing (cf. §3.3.1.2) proffered explanation on the part of advocates of the dynamical approach to SR cannot be applied in the GR context. Thus, for the advocate of the dynamical approach, there are two brute facts in GR—**MR1** and **MR2**—which lack further explanation from within the theory. Recall that these read as follows: (Cf. Ch. 1.)

MR1: All non-gravitational interactions are locally governed by Poincaré invariant dynamical laws.

MR2: The Poincaré symmetries of the dynamical laws governing non-gravitational fields in the neighbourhood of any point $p \in M$ coincide—in the regime in which terms featuring the Riemann tensor or its contractions may be ignored—with the symmetries of the dynamical metric field in that neighbourhood.

There are two points to make here. First, note that **MR1** held also in SR: that all non-gravitational interactions are (locally) governed by Poincaré invariant dynamical laws is a *brute fact*—an *outset assumption*—in both theories, which (the advocate of the dynamical approach contends) admits of no further explanation from within each theory. Second, as I argue in §3.5.3, while an untenable form of the geometrical approach may purport to account for both **MR1** and **MR2**, *any acceptable form of the geometrical approach must also accept these two miracles of GR*. In this sense, the existence of these two miracles is independent of the dynamical/geometrical debate.

3.4 Qualified and unqualified explanations

In §3.5, I consider whether a defensible version of the geometrical approach can be articulated. Before doing so, however, I first distinguish between what I call *qualified* versus *unqualified* explanations:

- (*Qualified explanations*.) Consider one particular dynamical equation featuring coupling to a metric field—for example, the special relativistic Klein-Gordon equation (3.2.1), or the general relativistic Klein-Gordon equation (3.2.2). Then ask: might the metric field in the theory in question (η_{ab} in the case of **KGS**; g_{ab} in the case of **KGG**) feature in an *explanation* of the form (in particular, of the symmetries) of that dynamical equation, and of the behaviour of the matter field(s) (here φ) to which it is coupled? Call this the question of *qualified explanation*—for the concern here is with accounting for the form of one, *given* dynamical equation, and for the behaviour of the particular fields coupled in that equation.

- (*Unqualified explanations.*) Consider *all possible* dynamical equations consistent with a given theory, such as SR or GR.[19] Then ask: does the metric field in the theory in question (η_{ab} in the case of SR; g_{ab} in the case of GR) explain the form (in particular, the symmetries) of *all those possible equations* consistent with the theory, and (in a certain particular way to be articulated) the behaviour of all possible matter fields, such that assumptions made in the formulation of the theory regarding the form of those equations and the behaviour of matter fields (e.g., that massless particles in GR traverse null geodesics) are, ultimately, redundant? For example, can η_{ab} explain the fact that all dynamical laws governing matter fields in SR are Poincaré invariant, or can g_{ab} in GR explain the SEP? Call this the question of *unqualified explanation*.

3.5 The geometrical approach

In this section, the above distinction between qualified and unqualified explanations is brought to bear on the question of whether there exists any viable form of the geometrical approach. My answer will be the following: while the form of the geometrical approach considered in e.g. [24, 28, 30] and Ch. 1 is *not* viable, there exists a weaker version of the approach, which *can* be defended.

The section proceeds as follows. In §3.5.1, I distinguish between these two versions of the geometrical approach, before exploring the different accounts they give

[19] There is some ambiguity regarding what is meant by a 'theory' here. To be clear, by 'theory' is meant here a theoretical *framework* such as that for SR or GR as presented in §3.2.3, rather than *specific* theories within those frameworks, such as **KGS** or **KGG**.

regarding the role of the metric field in explanations of the coincidence of (local) metric and dynamical symmetries, and of the behaviour of matter fields to which they couple, in both SR (§3.5.1.1) and GR (§3.5.1.2). In §3.5.2, I explore some further consequences of what I take to be the more defensible version of the geometrical approach. In §3.5.3, I demonstrate that this version of the geometrical approach does *not* account for **MR1** and **MR2**.

3.5.1 Two geometrical approaches

The geometrical approach, in both SR (§3.3.1.1) and GR (§3.3.2.1), may be understood in (at least) two different ways. Drawing upon the distinction presented in §3.4, the versions of the approach that I consider in this chapter are dubbed the *qualified* versus *unqualified* geometrical approaches:

- (*Qualified geometrical approach.*) Consider a particular dynamical equation governing the behaviour of a particular set of non-gravitational fields $\Phi_1, \ldots, \Phi_n$. Insofar as that equation features coupling to a metric field (as in e.g. (3.2.1) in **KGS**, or (3.2.2) in **KGG**), that metric field may contribute to an explanation of the symmetries of that dynamical equation, and of the dynamical behaviour of those $\Phi_1, \ldots, \Phi_n$ fields.

- (*Unqualified geometrical approach.*) Consider the metric field associated with a particular theory (for example, η_{ab} in SR, or g_{ab} in GR). That metric field constrains the form of all possible dynamical laws for non-gravitational fields consistent with that theory, such that assumptions about (local) dynamical symmetries are redundant in the formulation of the theory, and such that certain facts about the behaviour of matter fields are fixed.

In the following, I abbreviate 'the qualified geometrical approach' to **QGA**, and 'the unqualified geometrical approach' to **UGA**. On **QGA**, a particular metric field coupling to a particular set of non-gravitational fields in a particular dynamical equation may be understood to contribute to a qualified explanation (in the sense in §3.4) of the symmetries of that dynamical equation, and of the dynamical behaviour of those non-gravitational fields. On **UGA**, a particular metric field is taken to explain the symmetries of *all possible* dynamical equations in a given theory, and to fix certain facts about the behaviour of all possible matter fields, such that we need not, in fact, make any assumptions regarding dynamical symmetries, or about those dynamical facts, in that theory. (Importantly, I take both **QGA** and **UGA** to maintain the ontological autonomy of the metric field in both SR and GR.) It is principally **UGA** which is considered in Ch. 1 and attacked in [24,28,30], and it is this version of the geometrical approach which is (I maintain) untenable.

3.5.1.1 Special relativity

The reasons why **UGA** is untenable are similar in both the SR and GR cases; I begin with the former. The worry regarding **UGA** is put clearly by Brown and Pooley: [28, p. 84]

> As a matter of logic alone, if one postulates spacetime structure as a self-standing, autonomous element in one's theory, it need have no constraining role on the form of the laws governing the rest of the theory's models. So how is its influence supposed to work? Unless this question is answered, spacetime cannot be taken to explain the Lorentz covariance of the dynamical laws.

The point here is that it is consistent to have dynamical laws for non-gravitational fields in a theory featuring a Minkowski metric field η_{ab}, which nevertheless *do not* manifest the Poincaré symmetries of that metric field. As a concrete (trivial!) example, consider a modified version of **KGS**—call it **LAS**—KPMs of which are quadruples $\langle M, \eta_{ab}, \delta_{ab}, \varphi \rangle$, where δ_{ab} is a four-dimensional fixed *Euclidean* metric field,[20] and DPMs of which are picked out by the four-dimensional *Laplace* equation (hence my chosen nomenclature),[21]

$$\delta_{ab} \nabla^a \nabla^b \varphi = 0. \tag{3.5.1}$$

The dynamical symmetries of (3.5.1) do *not* include the Poincaré transformations (as for (3.2.1)); rather, they include the *Euclidean* transformations: those affine transformations the linear transformation matrix of which satisfies (cf. (3.2.5))

$$M_{\mu'}{}^{\mu} M_{\nu'}{}^{\nu} \delta_{\mu\nu} = \delta_{\mu'\nu'}. \tag{3.5.2}$$

LAS illustrates that a theory's featuring a certain metric field in its KPMs is *insufficient* for that theory's dynamical equations for non-gravitational fields to manifest the symmetries of that metric field, or for that metric field to play any constraining role in the dynamics of the matter fields in that theory, for those non-gravitational fields may couple to *other* fields (in this case, δ_{ab}), such that metric symmetries and dynamical symmetries do *not* coincide, and such that the matter fields manifest *other* dynamical behaviour (than that which they would manifest if they were coupled to the metric field under consideration, here η_{ab}). Of course, one may wish to exclude

[20]The notation δ_{ab} is chosen to emphasise the analogy with the Kronecker delta $\delta^a{}_b$; strictly, however, these are different objects, and should not be confused.

[21]Note that (3.5.1) is simply (3.2.1), with η_{ab} replaced by δ_{ab}; in making this move, the dynamical equation becomes an *elliptic*, rather than *hyperbolic*, partial differential equation.

coupling to such other fields; however, note that we then return to the situation in which the dynamical equations for matter fields manifesting certain symmetries, and yielding certain behaviour for those matter fields (e.g., that massless particles propagate on null geodesics), is an input *assumption*—it does not follow from (e.g.) η_{ab} alone.

On the other hand, **QGA** faces no such problems, for in this case the concern is not with generic, unqualified claims, but rather with the form of *one particular* dynamical equation and with the dynamical behaviour of the matter fields coupled in that equation. Why is (3.2.1) Poincaré invariant? Because it features coupling to the η_{ab} field—cf. §3.2.2. Why is (3.5.1) Euclidean invariant? Because it features coupling to δ_{ab}. Changing η_{ab} in (3.2.1) to δ_{ab} in (3.5.1) changes the behaviour of φ accordingly (after all, it is now governed by a different dynamical equation)—and it is very plausible to regard this as constituting a legitimate (if partial, for other factors may also be relevant to the dynamics of the field in question) *explanation* of the behaviour of φ. Thus, I take it that, in SR (and indeed, in the context of theories with fixed metric structure more generally), it is *incorrect* to regard as viable the explanation for the dynamical behaviour of matter proffered on the part of advocates of **UGA**, but *correct* to so regard the explanations proffered on the part of advocates of **QGA**. I discuss **QGA** further in §3.5.2.

3.5.1.2 General relativity

Similar points to those made above apply in the case of GR. According to advocates of **UGA**, the metric field g_{ab} in GR accounts for the local behaviour of all non-gravitational fields, such that the assumption of the SEP in the presentation of general relativistic theories in §3.2.3.2 is redundant, and such that matter fields must exhibit certain behaviour (e.g., such that test particles propagate on null geodesics). However, against such a claim, problem cases may also be identified.

One such example is the *Jacobson-Mattingly theory*, introduced in Ch. 1. Recall that the imposition of the Lagrange multiplier term in the action for this theory means that the dynamical behaviour of non-gravitational fields does *not* reflect the local (Poincaré) symmetries of the metric field. Rather, the (local) symmetries of the dynamical laws are a proper subset of the (local) metric symmetries. Given this, however, we appear to have in our possession a problem case for **UGA**, according to which the metric field *constrains* dynamical equations to manifest its own symmetries.

As with **LAS** in the case of SR, such cases appear to find their mark against **UGA**, for in the Jacobson-Mattingly theory, metric symmetries manifestly do *not* coincide with dynamical symmetries—so how could g_{ab} be constraining the local form of dynamical equations in this strong sense? On the other hand, **QGA** again does not appear to face such problems. For example, consider (3.2.2)—as in the SR context, it is perfectly reasonable to claim that the coupling in this equation of φ to g_{ab} offers an explanation of the dynamical behaviour of φ; moreover, the fact that no unqualified claim is made regarding possible form of dynamical equations for non-gravitational fields means that cases such as the Jacobson-Mattingly theory do not find their mark against **QGA** (for further discussion, see §3.5.2.2).

3.5.2 The qualified geometrical approach

I have argued that **QGA** is a defensible version of the geometrical approach, whereas **UGA** is not. In this subsection, I explore some further consequences of **QGA**. Specifically, I consider in §3.5.2.1 the sense in which the metric field in a given theory *may*, in fact, be understood to account for the form of *all* dynamical laws in that theory. In §3.5.2.2, I consider whether an account of the dynamical behaviour of matter in terms of metric structure is available on **QGA**, even in problematic cases such as those described above, in which (local) metric symmetries do not coincide with (local) dynamical symmetries.[22] I close in §3.5.2.3 by drawing a more fine-grained distinction within **QGA**.

3.5.2.1 Univocal explanation

In both SR and GR, there is a sense in which, on **QGA**, the metric field *can* explain the form of all dynamical laws in the theory—*once the restriction to a certain form of dynamical equations is made*. For example, *given* the restriction in SR to dynamical equations for non-gravitational fields which take a Poincaré invariant form, we may write all such equations in coordinate-free notion featuring coupling to η_{ab}[23]—in which case, η_{ab} may feature in explanations of the dynamical behaviour of the matter fields under consideration. This does not explain the *initial* restriction to Poincaré invariant dynamical laws for non-gravitational fields, but it *does* mean that η_{ab} may feature in explanations for the behaviour of all matter fields, *once such an assumption is made*. Similarly in GR, the metric field g_{ab} may not be able to account for the *initial* restriction to dynamical equations for matter fields obeying the SEP, but it *may* feature in explanations of the form of all dynamical laws for non-gravitational fields in GR, *once this assumption is made*—for in making this assumption, it is natural to consider dynamical equations in which matter fields are coupled to this very g_{ab} field.[24]

[22]Strictly, I will have to generalise the notion of a 'metric symmetry' in §3.5.2.2, to account for the examples given in that section. This, however, will be of no consequence.

[23]Cf. [30, §5].

[24]It is worth making two related points here. (1): Technically, such coupling is not essential, for we might instead couple to e.g. a fixed Minkowski metric field η_{ab}, or to a generic Lorentzian metric field which satisfies not the Einstein field equations, but some other set of dynamical equations. In the cases in which all dynamical laws feature coupling to g_{ab}, however, this metric field may feature in expla-

3.5.2.2 Partial explanation

A further subtlety regarding **QGA** pertains to the issue of *partial* explanation. I make the following claim: even in the cases in which metric and dynamical symmetries do not coincide, the metric field *may* feature in explanations of the dynamical behaviour of matter, on **QGA**. To see this, it is useful to consider three sub-cases: (i) situations in which dynamical symmetries form a proper subset of metric symmetries; (ii) situations in which dynamical symmetries form a proper superset of metric symmetries; (iii) cases where dynamical symmetries partially overlap with metric symmetries.

In order to discuss each of these cases, it is useful to introduce here three versions of Newtonian gravitation theory (NGT). First, let a *Leibnizian structure* be a triple $\langle M, t_{ab}, h^{ab} \rangle$, where M is a four-dimensional differentiable manifold; t_{ab} is a fixed temporal 'metric' field on M of signature $(1, 0, 0, 0)$; and h^{ab} is a fixed spatial (inverse) 'metric' field on M of signature $(0, 1, 1, 1)$.[25] The t_{ab} and h^{ab} fields are orthogonal, so that

$$h^{ab} t_{bc} = 0; \tag{3.5.3}$$

furthermore, I restrict in this chapter to structures (Leibnizian or otherwise; see below) which are *temporally orientable*, so that there exists a continuous (globally de-

nations of the form of all these laws. (2): One need not make the assumption that all dynamical laws manifest certain (local) symmetries so explicitly—one might instead make assumptions of (e.g.) *universal coupling* of the metric field to matter fields in *all* dynamical equations for the latter; this may, then, *entail* the relevant facts about the symmetries of those laws. This, indeed, appears to be Maudlin's stance, when he writes that "the fundamental requirement of a relativistic theory is that the physical laws should be specifiable using only the relativistic space-time geometry. For Special Relativity, this means in particular Minkowski space-time." [137, p. 117] The point here is that, on **QGA**, one may appeal to the metric field in giving certain generic explanations of the behaviour of matter fields in a certain restricted class of models of the theory—but the metric field itself does not account for those restrictions.

[25] Scare quotes are included on 'metric' here, for strictly neither t_{ab} nor h^{ab} satisfies the metric non-degeneracy condition—cf. [128, §4.1].

fined) one-form t_a that satisfies the decomposition condition $t_{ab} = t_a t_b$ at every point [128, p. 251].

In contrast with the notion of a Leibnizian structure, let a *Galilean structure* be a quadruple $\langle M, t_{ab}, h^{ab}, \nabla_a \rangle$, consisting of a Leibnizian structure, together with a derivative operator ∇_a on M satisfying the compaibility conditions

$$\nabla_a t_{bc} = 0, \tag{3.5.4}$$

$$\nabla_a h^{bc} = 0. \tag{3.5.5}$$

Finally, let a *Newtonian structure* be a tuple $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a \rangle$, consisting of a Galilean structure, together with a fixed vector field σ^a on M, such that

$$t_{ab}\sigma^b \neq 0. \tag{3.5.6}$$

Since none of Leibnizian, Galilean, or Newtonian structures are themselves metric fields, the notion of a metric symmetry cannot be applied in these cases.[26] However, the relevant notion easily generalises to the structures now under consideration: I say that a coordinate transformation is a *structure symmetry* just in case the structure under consideration is invariant under that transformation. Applying such a notion to Leibnizian, Galilean, and Newtonian structures, one finds that their associated structures symmetries are given by (no surprise!) the Leibniz, Galilean, and Newton groups.[27]

[26] For details regarding Leibnizian, Galilean, and Newtonian structures, see [64, ch. 2].

[27] The exact mathematical forms of these groups are not relevant for our purposes—see [169, §3.1] for details.

With these three structures in hand, we can consider three different theories—*viz.*, Newtonian gravitation theory set in each of these three structures. Consider first Newtonian mechanics set in a Galilean structure.[28] KPMs of this theory are tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \varphi, \rho \rangle$, where φ and ρ are real scalar fields on M, which will be taken to represent the gravitational potential and matter density, respectively. DPMs of this theory are picked out by the field equations[29]

$$R^a{}_{bcd} = 0, \tag{3.5.7}$$

$$h^{ab}\nabla_a\nabla_b\varphi = 4\pi\rho. \tag{3.5.8}$$

(3.5.7) imposes flatness of ∇_a; (3.5.8) is the Newton-Poisson equation. Finally, the gravitational force on a point (test) particle of mass m is given by $-mh^{ab}\nabla_b\varphi$; it follows from Newton's second law that, if this particle is subject to no forces except gravity, and given that it has four-velocity ξ^a, then it satisfies

$$-\nabla^a\varphi = \xi^b\nabla_b\xi^a. \tag{3.5.9}$$

Note that all elements of the Galilean structure feature in these dynamical equations; one can use this structure to offer a qualified explanation (in the sense of §3.4) of the form of these dynamical laws.

[28] A Galilean structure is traditionally considered to be the 'most appropriate' spacetime setting for NGT, for in this case structure symmetries and dynamical symmetries (are claimed to) coincide, thereby satisfying Earman's "adequacy conditions" on spacetime theories (see [64, §3.4]). For recent philosophical discussion calling into question whether this orthodoxy is correct, see [57, 112, 192, 206, 218, 224, 226]; I do not discuss further such matters in this thesis.

[29] Here, $R^a{}_{bcd}$ is the Riemann tensor associated with the derivative operator ∇_a defined in the Galilean structure.

Newtonian mechanics set in Galilean spacetime is a case in which structure symmetries coincide with dynamical symmetries.[30] Now consider a more nuanced case, in which dynamical symmetries constitute a proper subset of structure symmetries. One illustration of this is Newtonian mechanics set in a Leibnizian structure. KPMs of this 'theory' are tuples $\langle M, t_{ab}, h^{ab}, \varphi, \rho \rangle$ with $\langle M, t_{ab}, h^{ab} \rangle$ a Leibnizian structure, and φ and ρ defined as in the Galilean case; DPMs are (allegedly) picked out by (3.5.7)-(3.5.9). For the sake of argument granting that such a 'theory' is coherent,[31] we have a case in which dynamical symmetries are a proper subset of structure symmetries. What I contend here is that, in spite of the fact that structure symmetries and dynamical symmetries do not coincide, the fact that the Leibnizian structure still features in the DPMs of this theory means that it can still offer a *partial* (but not complete, since the laws also advert to other structure) explanation of the dynamical behaviour of matter in this case, in the qualified sense delineated in §3.4 above.

Next consider the case in which dynamical symmetries are a proper superset of structure symmetries.[32] An illustration of such a scenario is Newtonian mechanics set in a Newtonian structure. In this case, KPMs are tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho \rangle$, where $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a \rangle$ is a Newtonian structure (in which the integral curves of σ^a are taken to represent the worldlines of the persisting points of Newtonian absolute space), and φ and ρ are understood as above; DPMs of this theory are again picked out by (3.5.7)-(3.5.9). Though this theory is coherent,[33] as Earman states [64, §3.4], there is a sense in which it is nevertheless *malformed*, for the dynamical laws do not advert

[30]Setting aside the issues indicated in footnote 28.

[31]Indeed, I here include scare quotes on the word 'theory', as there are good grounds to question whether such a 'theory' is really coherent, since it does not have sufficient structure in its KPMs to be able to write down the dynamical equations used to fix its DPMs—cf. [198, p. 6]. (Belot puts the point pithily, when he accuses those working with such theories of "arrant knavery" [17, p. 571]; for further related discussion, cf. [57, pp. 268-269].)

[32]On this possibility, cf. Pooley's discussion at [169, p. 94].

[33]At least on **QGA**—it is questionable whether this theory is coherent on the dynamical approach, according to which (as discussed above) metric/structure symmetries in theories with fixed metric/structure (such as both SR and NGT) *just are* dynamical symmetries. Cf. [30, §3.1].

to all the Newtonian structure available in the KPMs of the theory (it is this which results in dynamical symmetries being a proper superset of structure symmetries). While I concur with Earman on this point, what I wish to register here is that, in this case, Newtonian structure *may* still be appealed to in explanations of the form of the dynamical laws governing matter fields—it is just that this structure has other, *redundant* explanatory apparatuses available to it (*viz*., the σ^a field).

Thus, on **QGA**, in the case in which dynamical symmetries are a subset of structure symmetries, the relevant structure (whether metric, or e.g. Leibnizian or Galilean or Newtonian) may feature in *partial* explanations of the dynamical behaviour of matter. In the case in which dynamical symmetries are a superset of structure symmetries, by contrast, the relevant structure may feature in *total* but *redundant* explanations of the dynamical behaviour of matter. Note that Jacobson-Mattingly theory instantiates the former case, in which dynamical symmetries are a subset of metric symmetries.[34]

Finally, consider the case in which dynamical symmetries partially overlap with structure symmetries—i.e., are neither a subset nor a superset of structure symmetries. One example of this is **LAS**, presented in §3.5.1.1. In this case, dynamical symmetries include the Euclidean transformations; symmetries of the Minkowski metric field η_{ab} are the Poincaré transformations. The intersection of the Euclidean and Poincaré groups is the group of translations and spatial rotations (cf. Ch. 1); therefore, the corresponding degrees of freedom associated to the η_{ab} field may still be used to account for *these* dynamical symmetries, in this case. (Though of course, an obvious question arises: why not instead appeal to δ_{ab} when giving this kind of qualified explanation of dynamical symmetries in this case?)

[34]Though in this case the theory is coherent, in a way that arguably NGT set in a Leibnizian structure is not—cf. footnote 31.

3.5.2.3 Confident and cautious qualified approaches

Suppose that one embraces **QGA**, and suppose that one is considering a theory in which the metric/structure under consideration can be appealed to in order to offer a qualified explanation of the symmetries of the dynamical laws governing matter fields. For example, suppose that one is considering theories such as **KGS**, or Newtonian mechanics set in a Galilean structure. Even in such cases, there exists a further question relevant to the chronogeometric significance of this metric/structure, on which one might take different views. Namely: do there actually exist rods and clocks which survey this metric/structure?

Different possible answers to this question distinguish two sub-views within **QGA**. On the one hand, one might maintain that, when the metric/structure features in a qualified explanation of the symmetries of the dynamical laws governing matter fields in the above sense, there always exist physical rods and clocks built from matter fields which survey that metric/structure. Call this view *confident* ***QGA***.[35] On the other hand, one might reject the claim that, when the metric/structure features in a qualified explanation of the symmetries of the dynamical laws governing matter fields in the above sense, there always exist physical rods and clocks built from matter fields which survey that metric/structure. Call this view *cautious* ***QGA***.

Clearly, in order to call into question confident **QGA**, it suffices to present a single problem case. In fact, there exist several such cases; here I mention two. First, Pitts presents in [162] the example of *universally coupled massive scalar gravity*. In such theories, there exist two Lorentizan metic fields: a dynamical field g_{ab}, and a fixed

[35] Arguably, Maudlin falls into this camp, for he both (a) speaks of restricting dynamical equations in SR to those which couple universally to η_{ab}, thereby placing him in **QGA** (cf. footnote 24); and (b) argues that, in any model of SR, there exists a clock which satisfies the *clock hypothesis*, and thereby (by definition) correctly reads off intervals along its worldline as given by the metric field (cf. [137, ch. 5]). There are good reasons to doubt (b)—cf. [140], discussed further below.

Minkowski metric field η_{ab}; the Lagrangian includes the following graviton mass piece:[36]

$$\mathcal{L}_{\text{mass}} = \frac{m^2}{64\pi G}\left[\frac{\sqrt{-g}}{w-1} + \frac{\sqrt{-g}^{w}\sqrt{-\eta}^{1-w}}{w\,(1-w)} - \frac{\sqrt{-\eta}}{w}\right] \tag{3.5.10}$$

(Here, w is a free parameter, which may be fixed to yield specific theories.) The important point to note about such theories is put clearly by Pitts: "Massive scalar gravity lacks Minkowskian behavior of rods and clocks, though it has the Minkowski metric (among other things) and the Poincaré symmetry group. ... [T]he chronogeometrically observable conformally flat metric $g_{ab} = \hat{\eta}_{ab}(-g)^{1/4}$ isn't clearly the One True Geometry [i.e., η_{ab}]."[37] [162, p. 6] Thus, theories of this kind appear to pose problems for confident **QGA**, for rods and clocks generally *do not* survey η_{ab}, in spite of the fact that this field couples to the matter fields in the theory, and so may feature in a qualified explanation of their symmetries.

As a second example, the authors of [140] (including the author of the present piece) demonstrate, drawing upon recent work by Asenjo and Hojman [6], that there should be no expectation that physical rods and clocks (such as light clocks) correctly survey the metric field g_{ab} of GR in particular solutions of this theory—namely in rotating solutions, such as the Gödel and Kerr solutions. The reasons are subtle, but essentially involve the fact that physical propagating media, such as light waves, do not travel at a fixed speed in such solutions, but rather manifest spacetime location-dependent propagation speeds. The central point here is a simple one: there is again reason to doubt confident **QGA**, for in these cases one has dynamical equations governing matter fields which feature coupling to g_{ab}, so that this metric field may feature

[36] For the full details, see [162].

[37] Indices in this passage have been altered for consistency with the present paper; there is no change in content.

in a qualified explanation of the symmetries of these equations and the behaviour of matter fields; nevertheless, rods and clocks do not survey this metric field, so that the chronogeometric significance of this field is questionable.

For these reasons, I take it that cautious **QGA** is to be preferred—no *a priori* assumptions should be made regarding the behaviour of physical rods and clocks, even in cases in which a partial explanation of (e.g.) the symmetries of the dynamical equations in the theory under consideration via a given metric/structure is possible. In the remainder of this paper, I set this distinction aside for simplicity—though (for the above reasons) it should be taken that reference to **QGA** always means reference to cautious **QGA**.

3.5.3 Two miracles, reprise

With these subtleties regarding **QGA** addressed, I close this section by arguing that this approach does *not* account for **MR1** and **MR2**; indeed, there is a sense in which **MR1** and **MR2** are *more* miraculous on **QGA**, than on the dynamical approach.

To see this, consider first SR on **QGA**. As in the case of the dynamical approach, on **QGA** it is conspiratorial—a 'miracle'—that all dynamical laws manifest the same symmetry properties, for recall that, unlike **UGA**, **QGA** seeks no explanation for this coincidence from within SR, in terms of η_{ab}. Put in other words, it is a *brute fact* on **QGA** that we do not consider other structures, such as δ_{ab}, to which the matter fields in the theory could couple, and as a result of which coupling their dynamical laws would manifest different symmetries. Thus, **MR1** holds also on **QGA**.

Since the advocate of **QGA** *also* considers even fixed metric structure such as η_{ab} to be ontologically autonomous, however, a second coincidence arises even in SR: why is it that the symmetries of this metric field coincide with the symmetries of all dynamical laws? Clearly, this is just **MR2**—again, another way to put the question is the following: why should the structure to which all dynamical laws for matter fields 'advert' be precisely the designated metric structure under consideration? From this, we see therefore that on **QGA**, both **MR1** and **MR2** hold even in the SR context. Since the dynamical approach faces the *single* miracle **MR1** in SR (since it ontologically reduces metric structure in this theory to dynamical symmetries), this is, arguably, reason to favour the dynamical approach over **QGA** in SR.

In the GR context, **QGA** also faces both **MR1** and **MR2**—for exactly the reasons delineated in §3.3.2.2. Given this, a new question arises: given that both the dynamical approach and **QGA** agree in the GR context that the g_{ab} field cannot be ontologically reduced to matter fields, and that both **MR1** and **MR2** hold in that context, is there really such a difference between the views, in this case? Absent the story of ontological reduction, there appears to be very little between the views. In light of this, I make the following claim: *While the dynamical approach and* ***QGA*** *are distinct in the context of theories with fixed metric structure such as SR (for they make different ontological claims regarding this fixed structure), they are not distinct in the context of theories with dynamical metric structure, such as GR.*[38]

[38]In this regard, cf. [169, p. 63], where Pooley writes, "What, then, is at stake between the metric-reifying relationalist and the traditional substantivalist? Both parties accept the existence of a substantival entity, whose structural properties are characterised mathematically by a pseudo-Riemannian metric field and whose connection to the behaviour of material rods and clocks depends on, *inter alia*, the truth of the strong equivalence principle. It is hard to resist the suspicion that this corner of the debate is becoming merely terminological."

3.6 The geodesic principle

So far, I have: (a) clarified the distinction between the dynamical and geometrical approaches—the latter itself coming in two distinct varieties: **UGA** and **QGA**; (b) argued that while **QGA** is viable, **UGA** is not; (c) demonstrated that **MR1** and **MR2** hold both on the dynamical approach and on **QGA**; (d) argued that there is no difference between **QGA** and the dynamical approach in the context of GR. In this section, I consider the connections between this work, and recent and important results on the *geodesic principle*. I also reflect upon work by Knox [111–113] and Weatherall [228, §6] pertinent to the themes of this chapter.

I begin with the geodesic principle. Contemporary work on this result stems largely from a 1975 theorem of Geroch and Jang [83]. Though more sophisticated extensions of this result now exist (in particular, see [67, 84]), I focus for the time being upon the Geroch-Jang theorem itself; this reads as follows:[39]

Theorem 1. (Geroch and Jang (1975)) *For a given $\langle M, g_{ab} \rangle$, where g_{ab} is a Lorentzian metric field on M, let $\gamma : I \to M$ be a smooth, embedded curve. Suppose that, given any open subset O of M containing $\gamma[I]$, there exists a smooth, symmetric field T^{ab} with the following properties:*

1. *T^{ab} satisfies the* strengthened dominant energy condition, *i.e. given any timelike vector ξ^a at any point $p \in M$, $T^{ab}\xi_a\xi_b \geq 0$ and either $T^{ab} = 0$ or $T^{ab}\xi_b$ is timelike;*
2. *T^{ab} satisfies the* conservation condition, *i.e.* $\nabla_a T^{ab} = 0$;
3. $\text{supp}\left(T^{ab}\right) \subset O$*; and*
4. *there is at least one point $p \in O$ for which $T^{ab}(p) \neq 0$.*

[39]Here, I use the notation of [228, p. 6].

Then γ is a timelike curve that may be reparameterised as a geodesic.

The Geroch-Jang theorem makes precise the essence of the geodesic principle: that small bodies move on geodesics. In [228, §6], Weatherall draws a number of philosophical lessons regarding geodesic theorems such as the above (and its more sophisticated successors), which he takes to be consonant with the dynamical approach; it is to these putative lessons that I now turn.[40] Begin with Weatherall's summary of the import of results such as the Geroch-Jang theorem:[41]

> [E]stablishing that small bodies respect the inertial structure encoded by a given derivative operator ∇_a requires one to establish that the T^{ab} field associated with matter is divergence-free, or "conserved", with respect to ∇_a.[42] [228, p. 36]

Weatherall takes the fact that T^{ab} is conserved with respect to a *specific* derivative operator ∇_a to deliver a connection between satisfaction of the geodesic principle and spacetime geometry—with this being particularly apparent if that derivative operator ∇_a is that which is compatible with some metric field: [228, p. 38]

> From this perspective it is also fair to say that, as Brown argues in *Physical*

[40]For Brown's own discussion of the geodesic principle, see [24, §9.3]. With Brown's central contention—that geodesic motion of small bodies in GR is a consequence of the Einstein field equations, and is therefore automatic in GR, in a way that it is not in antecedent theories ("It is no longer a miracle." [24, p. 163])—Weatherall is in disagreement, for (a) geodesic motion is, in fact, independent of the Einstein field equations; (b) similar results can be derived in other theories, e.g. NGT, and Newton-Cartan theory. (For the details of Newton-Cartan theory, in which the gravitational potential φ of NCT is absorbed into a (curved) derivative operator, see [128, ch. 4], or Ch. 6.) For Weatherall's work on the geodesic principle, see [220–222, 227, 228]; I am in agreement with him on these matters. Also worthy of mention in this regard are remarks in a similar vein to (a) made by Pooley [169, p. 543]; and an earlier paper of Malament [127], in which it is pointed out (*pace* Brown) that geodesic motion in GR follows *only* on the assumption of the strengthened dominant energy condition.

[41]Here, Weatherall's notation has been amended slightly: I use '∇_a' rather than '∇'.

[42]In addition to the satisfaction of the strengthened dominant energy condition—again, see the Geroch-Jang theorem as stated above.

> *Relativity*, spacetime structures such as the metric may be viewed as "a codification of certain key aspects of the behaviour of particles and fields" (p. 142), at least as regards the link between free, small-body motion and the privileged class of curves picked out by a metric and/or derivative operator.

Though I am in agreement with Weatherall as far as the above statements go, there remains more to be said here, on two fronts. First, though it is true that some connection between matter fields and geometry is forged insofar as the stress-energy tensor associated with these fields is conserved with respect to a specific derivative operator, and moreover insofar as that matter thereby follows geodesics of that derivative operator, in accordance with the Geroch-Jang theorem (or its extensions), thus far the connection proceeds in terms of the motion of *small bodies alone*. To move from such results regarding the geodesic motion of small bodies, to the behaviour of matter fields *tout court*, is in effect to demand that the local symmetry properties of all matter fields be derivable from such geodesic motions; that is, it is, in effect, to demand a proof of a result akin to *Schiff's conjecture*.[43] Only in that case could something like the SEP be delivered by this work on the geodesic principle.[44]

[43] In the words of Thorne *et al.*, "*Schiff's conjecture states that any complete and self-consistent gravitation theory that obeys [the weak equivalence principle] must also, unavoidably, obey [the strong equivalence principle]*" (emphasis in original) [207, p. 3575]. In turn, the weak equivalence principle is defined as follows: "*If an uncharged test body is placed at an initial event in spacetime, and is given an initial velocity there, then its subsequent worldline will be independent of its internal structure and composition*" (emphasis in original) [207, p. 3571]; the strong equivalence principle is defined as: "(i) [The weak equivalence principle] is valid, and (ii) the outcome of any local test experiment—gravitational or nongravitational—is independent of where and when in the universe it is performed, and independent of the velocity of the (freely falling) apparatus" [207, p. 3572]. For the original presentation of Schiff's conjecture, see [195, p. 343]; for ensuing discussion and attempted proofs of restricted versions of the conjecture, see [39, 126, 151, 207]. Clearly, the version of Schiff's conjecture under consideration in this chapter is different to that above—the gap to be bridged here is between the geodesic motions of small bodies, and the symmetries of matter fields *tout court*.

[44] Geroch and Weatherall demonstrate in [84] that source-free Maxwell fields 'track' null geodesics—a new result. Since the geodesic theorems demonstrate that massive matter moves on timelike geodesics, this gives access to both conformal and projective structure, respectively. One might think, therefore, that one may appeal to the Ehlers-Pirani-Schild result [68] (itself a generalisation of Weyl's

Second, it is important to be clear that this work does not provide a resolution to **MR1**. Even supposing that a connection is forged between geodesic motion and the local behaviour of matter fields more generally (*à la* Schiff's conjecture), that the mystery of **MR1** remains can be demonstrated through asking the following question: why should all matter fields have associated stress-energy tensors, the divergences of which vanish with respect to the *same* derivative operator? If this were not the case, then it need not be the case that all matter fields survey the same 'practical geometry', in the manner explicated by Weatherall. Though it is true that, as Weatherall observes [228, p. 11], the Einstein field equations tell us (via the contracted Bianchi identity) that the covariant divergence of the *total* stress-energy content of any particular solution of GR vanishes, this is (again, as Weatherall observes—see [228, p. 12]) insufficient to infer that the divergences of the stress-energy tensors associated with all *individual* matter fields vanish with respect to the same derivative operator. Thus, these results on the geodesic theorem do not place sufficient restrictions on the behaviour of even small bodies built from different matter fields in order to resolve **MR1**.

The situation regarding the bearing of these results upon **MR2** is more nuanced. Suppose that if the dynamical laws governing matter fields all manifest the same symmetries, then the stress-energy tensors associated with such matter fields (which satisfy the strengthened dominant energy condition, and the other conditions of the Geroch-Jang theorem and its generalisations) have covariant divergences which vanish with respect to the same derivative operator. Now suppose that the dynamical laws governing matter fields all manifest the same symmetries. Then (by the above), the stress-energy tensors associated with such matter fields have covariant

theorem—cf. [230]), that (subject to extra constraints) conformal and projective structure fixes metric structure, to strengthen the connection between these geodesic theorems and geometry. While such results do indeed yield a further sense in which local geometry may be inferred from geodesic motions, they continue to leave unbridged the gap between the geodesic motions of small bodies, and the local dynamics of matter *tout court*. That is, Schiff's conjecture remains unproven, in general.

divergences which vanish with respect to the same derivative operator. Then, the divergence of the total stress energy tensor (being a sum of the stress-energy tensors associated with the individual matter fields) with respect to this same derivative operator will also vanish; so, via the Einstein field equations, the left-hand side of the field equations will also have vanishing divergence with respect to this derivative operator—implying that the derivative operator is compatible with the metric field appearing in the Einstein tensor. In that case, small bodies built from all matter fields 'track' geodesics of a derivative operator associated with the Lorentzian metric field appearing in the Einstein field equations. In turn, one expects that in such a case the symmetries of the dynamical laws governing matter fields, and of this metric field, coincide, thereby delivering **MR2**. Of course, this reasoning is heuristic—but renders it *prima facie* plausible that these results regarding the geodesic principle may have application in resolving **MR2**.

In any case, let us now set aside these considerations regarding **MR1** and **MR2**, and focus upon Weatherall's general morals drawn in [228, §6]. Consider the following passage:

> [T]he reason that a metric (or metrics) and derivative operator are able to codify the behavior of (generic) matter in the way characterized by the geodesic principle is precisely that that metric and derivative operator are the ones that appear in the dynamics of (all) matter in the relevant ways. And this, I think, is ultimately what is at the heart of the matter.
>
> As I see it, the most perspicuous explication of what one means, or at least what one should mean, by the claim that spacetime has some geometry, represented by a given metric (or metrics) and derivative operator, is precisely that one can express the dynamics of (all) matter in such a way that all inner products are taken relative to that metric and all derivatives are

> taken relative to that derivative operator. This is the physical content of the claim that there are facts about distances, angles, and duration: physical processes occur in such a way that changes in a quantity at a time depend on the state of that quantity and those facts about distances, angles, and duration. And so, one is left with the conclusion that spacetime structures codify certain facts about the behavior of matter because the dynamics of (all) matter is adapted to those spacetime structures, which is just another way of saying that spacetime has that geometry. [228, pp. 39-40]

Though I am essentially in agreement with Weatherall on these matters, three points are important to note regarding this passage. First, and most straightforwardly, Weatherall (correctly) makes no appeal to **UGA**—he makes no claim to the effect that the metric field constrains all possible dynamical equations in a given theory, such that assumptions about the symmetry properties of those laws need not be made.

Second, nothing in this passage commits Weatherall either to the dynamical approach, or to **QGA**. Insofar as Weatherall takes e.g. NGT set in a Newtonian structure to be a coherent theory, there is perhaps some reason to take him to favour the latter, for recall that the coherence of this theory is questionable on the dynamical approach—cf. [30, §3.1].[45] Even in this case, however, one might take Weatherall's anticipated assessment that this theory is 'theoretically equivalent' (in a technical, category-theoretic sense—cf. [223, 225, 229]) to NGT set in a Galilean structure, combined with an implicit commitment to such theoretical equivalence being sufficient for physical equivalence, to indicate that he does *not* consider such to be the case—

[45] In more detail, recall from footnote 33 that, on the dynamical approach, metric/structure symmetries in theories with fixed metric/structure *just are* dynamical symmetries—so how could it be the case that there exists a theory in which such symmetries do not coincide?

meaning that perhaps he should be regarded as siding with advocates of the dynamical approach after all.[46]

Third, Weatherall's views as expressed in the above passage are very consonant with the 'spacetime functionalism' of Knox [111–113], according to which "the spacetime role is played by whatever defines a structure of local inertial frames" [113, p. 22] (cf. §3.1). To see this, some details regarding this programme of Knox must be recalled. Note first that, in GR, the chronogeometricity of the metric field precisely guarantees that this field be considered spatiotemporal, in Knox's sense. The reason is that, locally, the symmetries of the dynamical metric field coincide with those of the dynamical equations governing matter fields; in any frame in which these dynamical equations take their simplest form, the metric field itself takes the form $\mathrm{diag}\,(-1, 1, 1, 1)$. Thus, the metric field picks out a structure of local inertial frames—if one characterises such frames as those in which dynamical equations for non-gravitational fields take their simplest form (cf. [111, §2]).

Now recall that, for Weatherall, "what one means, or at least what one should mean, by the claim that spacetime has some geometry, represented by a given metric (or metrics) and derivative operator, is precisely that one can express the dynamics of (all) matter in such a way that all inner products are taken relative to that metric and all derivatives are taken relative to that derivative operator" [228, p. 40]. But, so coupling the dynamical equations governing matter fields will in general ensure that

[46] I concede that it is somewhat strained to seek to read Weatherall as an advocate of the dynamical approach; a reading on which he endorses something like **QGA** is more natural. Nevertheless, it is at least worth noting that advocation of the dynamical approach is *consistent* with Weatherall here. (Moreover—and interestingly—Weatherall has questioned in personal communication whether fixed metric structure, such as the Minkowski metric field of SR, should be regarded as being ontologically autonomous—in which case, his views are arguably closer to the dynamical approach than one might initially think. Whether, however, it is best to read Weatherall as endorsing the dynamical approach versus e.g. the version of the geometrical approach due to Janssen [12, 107, 108], in which the ontological autonomy of the metric field in e.g. SR is denied, remains unclear absent further work. Since the issues here are subtle, and it would take significant work to do justice to Janssen, these matters will have to wait for a future piece.)

those equations have certain local symmetry properties—as, for instance, our discussion of **KGG** illustrated. In particular, it will in general ensure that metric symmetries coincide (locally) with dynamical symmetries—that is, it will ensure that the metric field qualifies as spatiotemporal, on Knox's programme.[47] Thus, when Weatherall states that such coupling is sufficient for "spacetime to have some geometry", I take it that he is endorsing a view very much akin to Knox's spacetime functionalism.[48]

3.7 Conclusions

In the context of SR (and of theories with a fixed metric/structure[49] more generally), advocates of the dynamical approach maintain that such a metric/structure is ontologically reducible to (symmetries of the dynamical laws governing) non-gravitational fields. By contrast, in the context of GR (and of theories with a dynamical metric/structure more generally), no such claim is made on the part of advocates of the dynamical approach. As a result of this, the dynamical approach arguably collapses into **QGA** in the GR context. While the dynamical approach is distinct from **UGA** in both SR and GR, there are good reasons to doubt the plausibility of **UGA**.

On **QGA**, we can appeal to the metric field of e.g. SR or GR to explain certain universal facts about the dynamics of matter fields—but only once further restrictions on the allowed class of models under consideration are imposed (for example, assumptions regarding the symmetries of the dynamical laws for non-gravitational fields, or—relatedly—assumptions of the universal coupling of the metric field under

[47] This coupling will ensure that a necessary condition on the metric field's having chronogeometric significance is satisfied—cf. §3.3.

[48] Of course, it is also worth remaining conscious of the differences between Knox and Weatherall—for example, Weatherall makes no explicit commitment to inertial structure as the *sine qua non* of spacetime.

[49] 'Structure' construed here in the sense of §3.5.2.2.

consideration to the matter fields in those dynamical equations, etc.). Thus, on both the dynamical approach and **QGA**, as yet no complete account of **MR1** and **MR2** exists within GR. Indeed, while the dynamical approach faces only **MR1** in the context of SR, **QGA** faces both **MR1** *and* **MR2** in that theory; arguably, this reduction in the number of 'conspiracies' in SR constitutes reason to favour the former view over the latter. While work on geodesic principles establishes *some* connection between the dynamics of matter and the metric field of GR, this is in itself insufficient to account for **MR1**. Though there exist some hints that such results may be used to resolve **MR2**, more remains to be done in rendering these connections precise.

Weatherall may be understood as embracing Knox's spacetime functionalism, alongside either the dynamical approach, or **QGA**. Since both the dynamical approach and **QGA** are defensible, this is unproblematic. Indeed, advocates of the dynamical approach would do well to avoid writing off **QGA** too quickly—for this approach is perfectly viable, and makes no problematic *a priori* claim regarding the chronogeometric significance of the metric field in SR or GR, as with **UGA**. While I still incline to the dynamical view—essentially on grounds of ontological parsimony in theories such as SR—I hope this chapter may be of some value in demonstrating that the views of essentially all parties in this debate do not stand in such a state of conflict as one may *prima facie* be inclined to think.

Part II

Gravitational Energy

Chapter 4

Functional Gravitational Energy

Does the gravitational field described in general relativity possess genuine stress-energy? I answer this question in the affirmative, in (i) a weak sense applicable in a certain class of models of the theory, and (ii) arguably also in a strong sense, applicable in all models of the theory. In addition, I argue that one can be a realist about gravitational stress-energy in general relativity even if one is a relationist about spacetime ontology. In each case, my reasoning rests upon a functionalist approach to the definition of physical quantities.

4.1 Introduction

In the study of general relativity (GR), a range of claims have been advanced regarding the nature of stress-energy in the theory, and its status as a conserved quantity. For example, the intuition that stress-energy must be conserved in GR sometimes leads to the expression $\nabla_a T^{ab} = 0$ being treated as the conservation law for the stress-energy of matter fields (see, for example, [141, p. 152]); yet, on reflection, one may question whether this is a genuine conservation law at all. On the other hand, it is sometimes claimed that gravitational stress-energy exists in GR, but is non-local [91,141,196]. In this chapter, I argue that gravitational stress-energy does exist in the theory, in both (i) a weak sense applicable in a certain class of models of the theory, and (ii) arguably also in a strong sense, applicable in all models of the theory. In each case, my reasoning rests upon a functionalist approach to the definition of physical quantities.

In §4.2, I provide an interpretationally neutral presentation of the various stress-energy 'conservation laws' in GR. In §4.3, I argue for the existence of the above weak notion of gravitational stress-energy in certain models of GR, and also raise the possibility of a strong notion of gravitational stress-energy in the theory, applicable in all models of the theory. In §4.4, I consider ways in which the related debate between substantivalism and relationism on the ontology of spacetime connects with the debate on gravitational stress-energy in GR.

4.2 Dramatis Personæ

In this section, I introduce all mathematical concepts essential to foundational discussions of gravitational stress-energy in GR. In §4.2.1, I introduce the kinematically and dynamically possible models of GR; I then discuss in §4.2.2 differential and integral conservation laws in physical theories; before in §4.2.3 presenting specific stress-energy conservation laws which arise in special and general relativity. In §4.2.4, I introduce the mathematical apparatus of isometries and Killing vector fields, before finally discussing in §4.2.5 a conservation law for total (that is, matter-plus-gravitational) stress-energy in GR.

4.2.1 General relativity

Kinematically possible models (KPMs) of GR are picked out by triples $\langle M, g_{ab}, \Phi \rangle$, where M is a four-dimensional differentiable manifold;[1] g_{ab} is a Lorentzian metric field on M; and Φ is a placeholder for matter fields in the theory. Associated to $\langle M, g_{ab} \rangle$ there exists a unique derivative operator ∇_a, which is: (i) torsion free—in the sense that the associated torsion tensor $T^a_{\ bc}$, defined through $T^a_{\ bc} X^b Y^c = X^b \nabla_b Y^a - Y^b \nabla_b X^a - [X, Y]^a$, vanishes—and (ii) metric compatible—in the sense that $\nabla_a g_{bc} = 0$. Since this derivative operator ∇_a is uniquely determined by $\langle M, g_{ab} \rangle$, it is not included as an independent variable in the KPMs of GR.

Dynamically possible models (DPMs) of GR are those KPMs the geometrical objects of which satisfy the Einstein field equations[2]

$$G_{ab} = 8\pi T_{ab} \tag{4.2.1}$$

—the dynamical equations of the theory, which relate g_{ab} to the stress-energy tensor T_{ab} of the Φ—in addition to the dynamical equations of the Φ.

[1]One should avoid, at this stage, asserting M to be the 'spacetime manifold', for to do so is to conflate the mathematical model under consideration with the possible world to which that model is ultimately interpreted as corresponding. Indeed, in light of the debate over the hole argument [65], it is not necessarily correct to interpret M as representing substantial spacetime—see §4.4.1.

[2]Until §4.4, I restrict to the sector of GR with vanishing cosmological constant Λ. For $\Lambda \neq 0$, (4.2.1) reads $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$.

4.2.2 Differential and integral conservation laws

4.2.2.1 Case 1: flat connection

For a given derivative operator ∇_a and tensor field $T^{a_1 \ldots a_n}$, suppose

$$\nabla_{a_1} T^{a_1 \ldots a_n} = 0. \tag{4.2.2}$$

In a coordinate basis, (4.2.2) reads

$$\partial_{\mu_1} T^{\mu_1 \ldots \mu_n} + \Gamma^{\mu_1}{}_{\sigma\mu_1} T^{\sigma\mu_2 \ldots \mu_n} + \cdots + \Gamma^{\mu_n}{}_{\sigma\mu_1} T^{\mu_1 \ldots \mu_{n-1}\sigma} =: \partial_{\mu_1} T^{\mu_1 \ldots \mu_n} + \Delta^{\mu_2 \ldots \mu_n} = 0, \tag{4.2.3}$$

where the $\Gamma^{\mu}{}_{\nu\sigma}$ are the connection components associated to the derivative operator ∇_a in this coordinate basis,[3] and I have introduced the notation $\Delta^{\mu_2 \cdots \mu_n}$ to represent the terms featuring connection components in the above expression. In GR,[4] at any $p \in M$ we can find Riemann normal coordinates such that connection components vanish, and (4.2.3) becomes

$$\partial_{\mu_1} T^{\mu_1 \cdots \mu_n}(p) = 0. \tag{4.2.4}$$

This results holds only at p. If we wish (4.2.4) to obtain at some other $q \in M$ in the neighbourhood of p, we require that derivatives of connection components vanish—

[3] In a coordinate basis $\{e_\mu\}$, the connection components are defined through $\nabla_\rho e_\nu =: \Gamma^{\mu}{}_{\nu\rho} e_\mu$.
[4] So using the derivative operator ∇_a introduced in §4.2.1.

and *a fortiori* that the derivative operator ∇_a be flat.[5] In this case, we may write

$$\partial_{\mu_1} T^{\mu_1 \cdots \mu_n} = 0. \tag{Par-T}$$

Given the local validity of (Par-T), we may integrate over a volume $V \subset M$ and apply Gauss' theorem, to obtain (n^a is a unit normal vector to $S = \partial V$; $n_a = g_{ab} n^b$)

$$\int_{S=\partial V} n_{\mu_1} T^{\mu_1 \cdots \mu_n} dS = 0. \tag{Int-T}$$

Note that (Par-T) is invariant only under affine transformations—that is, only retains the same form in coordinate systems related by such transformations.[6] This means that an equation of the form (Int-T) also holds only in such coordinate systems. If one applies a non-affine transformation $x^{\mu} \rightarrow x^{\mu'} = f^{\mu'}(x^{\mu})$ to (Par-T),[7] one will generically pick up extra terms (which we may schematically represent by $\Theta^{\mu_2 \cdots \mu_n}$), so that this equation reads instead[8]

$$\partial_{\mu_1'} T^{\mu_1' \cdots \mu_n'} + \Theta^{\mu_2' \cdots \mu_n'} = 0. \tag{4.2.5}$$

[5] Since in a coordinate basis, the (unique) Riemann tensor $R^a{}_{bcd}$ associated to ∇_a—defined through $R^a{}_{bcd}\xi^b = -2\nabla_{[c}\nabla_{d]}\xi^a$ for all smooth fields ξ^a [128, p. 68]—reads $R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} - \partial_\sigma \Gamma^{\mu}{}_{\nu\rho} + \Gamma^{\tau}{}_{\nu\sigma}\Gamma^{\mu}{}_{\tau\rho} - \Gamma^{\tau}{}_{\nu\rho}\Gamma^{\mu}{}_{\tau\sigma}$.

[6] Consider an affine coordinate transformation $x^{\mu'} = M^{\mu'}{}_{\mu} x^{\mu} + a^{\mu'}$. If an (r,s) tensor field $F^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s}$ transforms under this coordinate change as $F^{\mu_1 \cdots \mu_r}{}_{\nu_1 \ldots \nu_s} \rightarrow M^{\mu_1}{}_{\mu_1'} \ldots M^{\mu_r}{}_{\mu_r'} M^{\nu_1'}{}_{\nu_1} \ldots M^{\nu_s'}{}_{\nu_s} F^{\mu_1' \cdots \mu_r'}{}_{\nu_1' \ldots \nu_s'}$, then we say that it is 'covariant' with this coordinate transformation. If a dynamical equation retains the same form in either of the two coordinate systems under consideration, then we say that it is 'invariant' under the coordinate change.

[7] The notation $x^{\mu'} = f^{\mu'}(x^{\mu})$ signifies that in this case $x^{\mu'}$ may be defined in terms of arbitrary contractions with x^{μ}, provided that there is one free primed index.

[8] To take an explicit example (relevant to our discussions of stress-energy below), consider the expression $\partial_\mu T^{\mu\nu} = 0$. Transforming to a new coordinate basis $x^{\mu} \rightarrow x^{\mu'} = f^{\mu'}(x^{\mu})$, one obtains $\partial_{\mu'} T^{\mu'\nu'} + \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial^2 x^{\mu}}{\partial x^{\mu'} x^{\lambda'}} T^{\lambda'\nu'} + \frac{\partial x^{\nu'}}{\partial x^{\nu}} \frac{\partial^2 x^{\nu}}{\partial x^{\lambda'} \partial x^{\sigma'}} T^{\lambda'\sigma'} =: \partial_{\mu'} T^{\mu'\nu'} + \Theta^{\nu'} = 0.$

Integrating over the volume V as before, we have by analogy with (Int-T),

$$\int_{S=\partial V} n_{\mu_1'} T^{\mu_1' \cdots \mu_n'} dS = -\int_V \Theta^{\mu_2' \cdots \mu_n'} dV. \tag{4.2.6}$$

4.2.2.2 Case 2: non-flat connection

Consider now the case in which ∇_a is not flat. In that scenario, (Par-T) does not hold in the neighbourhood of p,[9] and one may not integrate to construct a result of the form (Int-T). Instead, one has only the more general equation (4.2.3); integrating this expression and applying Gauss' theorem, one obtains

$$\int_S n_{\mu_1} T^{\mu_1 \cdots \mu_n} dS = -\int_V \Delta^{\mu_2 \cdots \mu_n} dV. \tag{4.2.7}$$

In a new coordinate basis $x^{\mu'} = f^{\mu'}(x^\mu)$, (4.2.3) transforms to the same expression written in primed indices;[10] integrating and applying Gauss' theorem, one obtains (4.2.7) written in primed indices. Since the right-hand side of (4.2.7) is coordinate-dependent (in the sense that it consists of non-tensorial geometrical objects), in general the value of the integral on the left-hand side of (4.2.7) will differ when this expression is written in unprimed versus primed coordinates.

From these results, one concludes the following. If the derivative operator ∇_a is flat, then an equation of the form (Par-T) holds in a class of frames related by affine transformations;[11] given this, an equation of the form (Int-T) also holds in such frames. In a generic coordinate system at $p \in M$, however, such results do not hold

[9] Rather, (Par-T) holds only at p.

[10] This result is straightforward: (4.2.3) follows from (4.2.2), which involves only tensorial quantities.

[11] Defined through specifying normal coordinates at $p \in M$.

(even when ∇_a is flat)—one has, rather, (4.2.5) and (4.2.6). If the derivative operator in question is not flat, then in general equations of the form (Par-T) or (Int-T) do not hold; one has instead equations of the form (4.2.3) and (4.2.7). Since the right-hand side of (4.2.7) is coordinate-dependent (in the sense of being composed of non-tensorial quantities), so too is the value of the integral on the left-hand side.

4.2.2.3 Conservation laws

In physics, one often speaks both of differential and integral conservation laws—the former of form (Par-T); the latter of form (Int-T). To illustrate why these are referred to as 'conservation laws', substitute the second rank tensor $\left(T^{\mu\nu}_{\Phi_1} + T^{\mu\nu}_{\Phi_2}\right)$ for $T^{\mu\nu_1\ldots\nu_i}$—with $T^{\mu\nu}_{\Phi_i}$ here interpreted as the stress-energy tensor of the ith matter field (in the coordinate basis under consideration). In this case, the above two laws become $\partial_\mu\left(T^{\mu\nu}_{\Phi_1} + T^{\mu\nu}_{\Phi_2}\right) = 0$ and $\int_S n_\mu \left(T^{\mu\nu}_{\Phi_1} + T^{\mu\nu}_{\Phi_2}\right) dS = 0$, respectively. Trivially, moving the terms for Φ_2 to the right hand side, we obtain for (Par-T)

$$\partial_\mu T^{\mu\nu}_{\Phi_1} = -\partial_\mu T^{\mu\nu}_{\Phi_2}, \tag{4.2.8}$$

and for (Int-T)

$$\int_S n_\mu T^{\mu\nu}_{\Phi_1} dS = -\int_S n_\mu T^{\mu\nu}_{\Phi_2} dS. \tag{4.2.9}$$

(4.2.8) tells us that, at any point, any change in the stress-energy of the Φ_1 field must be balanced by an equal and opposite change in the stress-energy of the Φ_2 field; stress-energy is therefore conserved between Φ_1 and Φ_2 at that point. Analogously, (4.2.9) tells us that any change in the stress-energy of Φ_1 across a boundary of a spacetime region $S \subset M$ must be balanced by an equal and opposite change in the stress-energy of Φ_2 across that boundary; stress-energy is therefore conserved in the volume enclosed by S.

Prima facie, both (Par-T) and (Int-T) are legitimate conservation laws. In a sense though, the differential conservation law (Par-T) is stronger than the integral conservation law (Int-T), as while the former implies the latter, the reverse is not true. Though I return to this point in §4.3.2, I emphasise at this stage that there is no reason to think that (Par-T) and (Int-T) are not both legitimate conservation laws, albeit different in scope.

4.2.3 Conservation equations in special and general relativity

4.2.3.1 Special relativity

The above in hand, consider the role of stress-energy conservation laws in the context of relativity theory. In special relativity (SR), the 'differential conservation law' for the stress-energy tensor of matter fields T^{ab} reads

$$\nabla_a T^{ab} = 0, \qquad \text{(Cov-T)}$$

where ∇_a is the unique torsion-free derivative operator compatible with the fixed Minkowski metric field η_{ab} of SR.[12] Crucially, this derivative operator is flat—meaning that the discussion of §4.2.2 applies, and in SR we have a differential conservation law for the stress-energy tensor of matter fields of the form (Par-T) in the neighbour-

[12] As discussed by Pooley [171, p. 115], and unlike the case of the metric field g_{ab} of GR, one may understand η_{ab} as being fixed identically in all KPMs of SR. Otherwise, KPMs of SR—as with GR—are again denoted by triples, this time of the form $\langle M, \eta_{ab}, \Phi\rangle$.

hood of a given $p \in M$, when (Cov-T) is written in normal coordinates at p, for all frames related to such coordinates by affine transformations. By integrating, we obtain a well-defined 'integral conservation law' of the form (Int-T)—in the same class of frames in which the associated 'differential conservation law' holds. When the above conditions are *not* satisfied, (Cov-T) instead takes the form (4.2.5).

4.2.3.2 General relativity

In GR, the fixed metric field η_{ab} of SR is replaced with a generic Lorentizan metric field g_{ab}, and the derivative operator ∇_a is now the unique torsion-free such operator compatible with g_{ab}.[13] In this case, ∇_a can no longer be assumed to be flat; moreover, the operator is rendered dynamical, due to the coupling of curvature and matter degrees of freedom via (4.2.1). Given this, in GR, (Cov-T) yields 'conservation laws' of the form (4.2.3) and (4.2.7).

4.2.3.3 Discussion

Though many textbooks claim that (Cov-T)—subject to its GR interpretation—is the differential conservation law for matter stress-energy in GR (see, for example, [141, 196,214]), foundational authors often claim to the contrary that this is not a legitimate conservation law (see, for example, [7,95,118,155]). To assess such claims, first consider the form (4.2.3) to which (Cov-T), subject to its GR interpretation, reduces in a coordinate basis. Since (4.2.3) does not take the form (Par-T), it cannot *prima facie* be understood as a differential conservation law in the sense of §4.2.2.3 [7,95,118]. In fact, a common interpretation of (Cov-T), subject to its GR interpretation and writ-

[13] Performing these two replacements in any dynamical equations of SR, to obtain general relativistic dynamical equations, is sometimes dubbed 'minimal coupling'. For philosophical discussion, see Ch. 1.

ten in a coordinate basis, is that locally no rest mass or momentum is created, except by the interaction between the gravitational field and matter fields. Indeed, for this reason Ohanian states '[(4.2.3)] is really a *non-conservation* law!' [156, p. 3]. While it is true that (4.2.3) does not take the canonical form (Par-T), there are two immediate problems with this interpretation:

1. To claim that $\Delta^{\mu_1 \cdots \mu_n}$ terms in equations such as (4.2.3) represent interactions between the gravitational and matter fields is to overlook a debate in the history and philosophy of GR as to which mathematical object should be identified with the gravitational field. Specifically, while some argue that the metric field g_{ab} represents the gravitational field [120, §4.3], others (arguably including Einstein [77][14]) argue that the connection coefficients $\Gamma^{\mu}{}_{\lambda\nu}$ (in any given coordinate basis) play this role, and still others argue that the gravitational field is represented by the Riemann tensor $R^{a}{}_{bcd}$ [203, p. 8].[15] Though ultimately one may be able to argue that the gravitational field is best represented by the connection coefficients $\Gamma^{\mu}{}_{\lambda\nu}$ (a point to which we return shortly), to simply assert this is to overlook an important debate in the foundations of GR.

2. To claim that $\Delta^{\mu_1 \cdots \mu_n}$ terms in (4.2.3) denote interactions between the gravitational and matter fields is to conflate the Φ in a given KPM of GR with the associated stress-energy tensor T_{ab}. These objects are not the same, so saying that such terms represent the interaction of matter with the gravitational field must be accompanied with a precise explication of the sense in which this is so.

We return to the first of these points in the forthcoming sections. For now, it suf-

[14]To claim that Einstein identified the gravitational field with the connection coefficients may be to oversimplify his position. In fact, Einstein saw GR as unifying gravity and inertia, in the same way that SR had unified electricity and magnetism. If one takes this view, then perhaps one need never speak of the gravitational field in GR. See [29, 122].

[15]For further discussion of these matters, see [120].

fices to note that the fact that (4.2.3) contains extra terms which result in it not taking the form (Par-T) results in difficulties in interpreting this as a standard differential conservation law.

While the problems with treating (Cov-T) subject to its GR interpretation as a conservation law are widely discussed in the literature, the analogous case of (Cov-T) subject to its special relativistic interpretation also merits consideration. In that case, in an arbitrary frame of reference in the neighbourhood of a point $p \in M$, the very same points as outlined above apply, for in such a case the differential and integral 'conservation laws' for the theory become, respectively, (4.2.5) and (4.2.6). Thus, one can say that it is only in particular frames of reference—those related to a choice of normal coordinates at p by affine transformations—that one can construct uncontroversial differential and integral conservation laws (respectively (Par-T) and (Int-T)) in SR. In the following section I introduce some further mathematical apparatus, which will enable me to shed light upon this result.

4.2.4 Killing vector fields and spacetime isometries

A diffeomorphism $\phi : M \to M$ is said to be a 'symmetry transformation' of a tensor field T just in case $\phi_* T = T$, where $\phi_* T$ is the push-forward of T. A symmetry transformation of the metric field is called an 'isometry'. Such isometries can be characterised by their generators—that is, their associated Killing vector fields—defined through Killing's equation,[16]

$$\nabla_{(a}K_{b)} = 0. \tag{4.2.10}$$

To every such Killing vector field there corresponds an integral stress-energy conservation law. To see this, note that if a given $\langle M, g_{ab}\rangle$ possesses a Killing vector field K^a, then one can build the quantity $T^{ab}K_b$ with vanishing divergence:

$$\nabla_a\left(T^{ab}K_b\right) = \nabla_a T^{ab}K_b + T^{ab}\nabla_a K_b = T^{ab}\nabla_{(a}K_{b)} = 0. \tag{4.2.11}$$

In the penultimate step here, I have used the symmetry of T^{ab}; in the final step I have used (4.2.10). To see how this leads to a well-defined integral conservation law even in the case in which ∇_a is not flat, now recall the covariant divergence theorem,

$$\int_M d^n x\sqrt{|g|}\nabla_a X^a = \int_{\partial M} d^{n-1}x\sqrt{|h|}n_a X^a, \tag{4.2.12}$$

which holds whenever ∂M is timelike or spacelike; X^a is a vector field on M; ∇_a is the unique torsion-free derivative operator compatible with g_{ab}; g denotes the determinant of g_{ab}; and h denotes the determinant of the metric on ∂M induced by pulling

[16]The 'Lie derivative' $\mathcal{L}_X T$ of a tensor field T represents how that tensor field changes as one acts with a diffeomorphism along the integral curves of a vector field X^a; the condition $\phi_* T = T$ imposes that $\mathcal{L}_X T = 0$. Since the Lie derivative of the metric field reads $\left(\mathcal{L}_X g\right)_{ab} = 2\nabla_{(a}X_{b)}$ (assuming that the unique metric-compatible, torsion free derivative operator ∇_a is used), this condition yields $\nabla_{(a}X_{b)} = 0$, which is Killing's equation. See [214, pp. 437-444]. Note also that $K_a = g_{ab}K^b$.

back the metric on M.[17] Using (4.2.12), one may integrate (4.2.11) to obtain

$$\int_V \nabla_a T^{ab} K_b dV = \int_{S=\partial V} n_a T^{ab} K_b dS = 0, \tag{4.2.13}$$

where n^a is the unit normal to S ($n_a = g_{ab} n^b$), and dS and dV are respectively surface and volume elements. This integral formulation unequivocally means (non-gravitational) stress-energy conservation, with respect to the integral curves of the Killing vector field K^a. (4.2.13) shows that we can have (non-gravitational) stress-energy integral conservation laws in curved spacetime in cases where this latter instantiates certain global symmetries; namely, when spacetime structure remains stationary along the integral curves of a Killing vector field.[18,19]

How does this relate to the work undertaken above? To see the connection, first recall that a special relativistic 'spacetime' $\langle M, \eta_{ab} \rangle$ has ten independent Killing vector fields; these correspond to the generators of Poincaré transformations.[20] Thus, if one projects $\nabla_a T^{ab}$ onto such a Killing vector K^a as in (4.2.11) at every point in

[17]See [214, pp. 433-434].

[18]This does not preclude the possibility of other unspecified ways of obtaining genuine (non-gravitational) stress-energy conservation in curved spacetime: it specifies sufficient but not necessary conditions.

[19]The covariant divergence theorem can only be applied to vector fields, but not to higher rank tensor fields: this is why we obtained in (4.2.13) an integral conservation law from $\nabla_a \left(T^{ab} K_b \right) = 0$, but we cannot obtain an analogous integral conservation law from $\nabla_a T^{ab} = 0$.

[20]Consider a Killing vector field K^a associated to $\langle M, g_{ab} \rangle$, which by definition satisfies (4.2.10) (the derivative operator ∇_a in this equation is the unique torsion-free such operator compatible with g_{ab}, as usual). From this, one can derive straightforwardly that K^a must also satisfy $\nabla_a \nabla_b K^c = R^c{}_{bad} K^d$. Restricting to the special relativistic case $\langle M, \eta_{ab} \rangle$, one has $R^c{}_{bad} = 0$, in which case $\nabla_a \nabla_b K^c = 0$. Then, restricting to normal coordinates at some $p \in M$, one has that $\partial_\mu \partial_\nu K_\rho = 0$—and integrating this equation, one finds that $K_\mu = a_{\mu\nu} x^\nu + b_\mu$ where b_μ is a constant one-form, and the coefficients of $a_{\mu\nu}$ are also constant. Therefore, the components K_μ are linear functions of the inertial frame coordinates. Now, substituting this result into (4.2.10) reveals that $a_{\mu\nu}$ must be antisymmetric—i.e. has six independent components. Since b_μ has four independent components, in total we find that there are ten independent Killing vector fields associated to the 'spacetime' $\langle M, \eta_{ab} \rangle$; these are the isometries of the Minkowski metric η_{ab}. Since the Poincaré group is the group of coordinate transformations which leave the Minkowski metric invariant, we see that this group must have ten independent generators, each corresponding to a Killing vector field.

the neighbourhood of some $p \in M$, one projects onto the integral curves of a vector field which generates Poincaré transformations—and so along which the conservation laws (Par-T) and (Int-T) hold (see §4.2.2). By using the covariant divergence theorem, one is then able to construct an integral conservation law (4.2.13) which holds in any arbitrary frame.

4.2.5 The gravitational stress-energy pseudotensor

We have seen in §4.2.2.1 that in SR, a conservation law of the form (Par-T) holds in a class of frames related by Poincaré transformations (in fact, all affine transformations); this can be expressed in covariant language through contraction with the Killing vector fields K^a associated to the isometries of $\langle M, \eta_{ab} \rangle$, as per (4.2.13). In GR, by contrast, no such move is possible in general (see §4.2.2.2)—as can be seen through the fact that in general a 'spacetime' $\langle M, g_{ab} \rangle$ satisfying (4.2.1) possesses no non-trivial Killing vector fields. This notwithstanding, however, one might seek to construct an alternative stress-energy conservation principle in GR.

One way to reconcile one's intuition that stress-energy must be conserved in GR with the observation that (Cov-T)—subject to its GR interpretation—is not a conservation law of the form (Par-T) is to argue that there instead exists a conservation law of the form (Par-T) in models of the theory for material plus 'gravitational' stress-energy (represented in a given coordinate basis by $t^{\mu\nu}$[21]):

$$\partial_\mu \mathcal{T}^{\mu\nu} = \partial_\mu \left(T^{\mu\nu} + t^{\mu\nu} \right) = 0. \tag{Par-Tt}$$

The idea is that, though generically $T^{\mu\nu}$ alone is not conserved (in every frame, at some $p \in M$) in the sense of (Par-T) in GR, perhaps a quantity representing the stress-energy of the gravitational field exists such that the sum of the two is a conserved quantity (that is, such that (Par-Tt) holds in every frame, at a given $p \in M$). The vanishing of $\partial_\mu \mathcal{T}^{\mu\nu}$ can be encoded in terms of an antisymmetric 'superpotential' $U^{\mu\lambda\nu} = U^{\mu[\lambda\nu]}$, by writing (see, for example, [118, 209])

$$T^{\mu\nu} + t^{\mu\nu} = \partial_\lambda U^{\mu\lambda\nu}. \tag{4.2.14}$$

Bearing in mind (4.2.1) and (4.2.14), one may choose to define $t^{\mu\nu}$ (in a given frame) via

$$t^{\mu\nu} := \partial_\lambda U^{\mu\lambda\nu} - \frac{1}{8\pi} G^{\mu\nu}. \tag{4.2.15}$$

In fact, there is a freedom of choice of the superpotential, since (Par-Tt) does not specify this object uniquely. This leads to distinct, non-equivalent expressions for $t^{\mu\nu}$,

[21] Three notes on $t_{\mu\nu}$ are in order. (1) This term and the notation used to denote it were originally introduced by Einstein in [71, 78]. Einstein went on to state in 1918 that "nearly all my colleagues raise objections to my definition of the momentum-energy theorem" [74, p. 448]; here he had in mind particularly Levi-Civita [124], Schrödinger [193], and Bauer [13], with whom he had corresponded heavily on this topic in the preceding four years. (2) Some, such as Hoefer [95, p. 193] and Nerlich [149, p. 162], have taken $t^{\mu\nu}$ to be defined implicitly through (4.2.3). Unfortunately, this does not work out smoothly, as there is more than one candidate for $t^{\mu\nu}$ (as discussed below), so (4.2.3) is not sufficient to fix $t^{\mu\nu}$ uniquely. (3) For simplicity in this chapter I use the symbol $t^{\mu\nu}$ to refer both to the object representing gravitational stress-energy, and to its components in a given coordinate basis.

including (among others) the so-called 'Einstein pseudotensor' and 'Landau-Lifschitz pseudotensor'.[22] For my purposes, it is crucial to note that $t^{\mu\nu}$ does not transform as a tensor under a coordinate change; it is for this reason that it is often referred to as the gravitational stress-energy 'pseudotensor'. In particular, at any point $p \in M$ there is a coordinate system in which $t^{\mu\nu}$ vanishes.

It is worth emphasising that what I am referring to as 'the' gravitational stress-energy psuedotensor is doubly ambiguous, in the following sense:

A. There are many distinct but non-equivalent choices for this pseudotensor, based upon one's choice of superpotential. Hence, when one refers to 'the' gravitational stress-energy pseudotensor, one is implicitly supposing that a choice has been made from the family of possible candidates.

B. Once one such definition of this pseudotensor is chosen, the resulting object is still a frame-dependent (that is, non-tensorial) quantity.

With these points regarding the gravitational stress-energy pseudotensor in hand, I consider in §4.3 possible interpretations of (Par-Tt). Before doing so, however, one further observation is in order: the requirement that there be a conservation principle of the form (Par-Tt) in GR which holds in every frame is *prima facie* a very strong condition—for typically one does not consider such a conservation principle to hold even in SR (see §4.2.3.1). That said, she who requires (Par-Tt) to hold in every frame in GR could (in principle) define an analogous principle to hold in every frame in SR. This issue is discussed further in §4.3.

[22] There exist interesting questions regarding which of these non-equivalent versions of the gravitational stress-energy pseudotensor best describes gravitational stress-energy. See [209, pp. 190-191].

4.3 Interpreting Conservation Laws for Total Stress-Energy

In this section, I assess whether putative conservation laws for total (i.e. matter-plus-gravitational) stress-energy in GR such as (Par-Tt) can indeed be regarded as legitimate conservation principles, and whether there is any physical significance to the notion of 'gravitational stress-energy' in models of GR. To this end, in §4.3.1 I consider possible interpretations of (Par-Tt), finding that one's verdict on whether this equation counts as a conservation law for total stress-energy in GR depends upon one's view on whether non-tensorial objects such as $t^{\mu\nu}$ may represent physical quantities. In §4.3.2, I use (Par-Tt) to construct an integral conservation law, initially drawing similar conclusions. However, I then show that in certain physical circumstances, this integral version of (Par-Tt) yields a notion of gravitational stress-energy at least as robust as that in SR. In §4.3.3, I reflect on the correct attitude that one should take towards this quantity, and therefore on whether one should be a realist about gravitational stress-energy in GR in this sense.

4.3.1 The differential conservation law for total stress-energy

Does (Par-Tt) qualify as a conservation law for total stress-energy in GR? One's answer to this question hinges upon one's understanding of which mathematical objects in a theory should be taken to represent physical entities. In modern works on GR, something like the following is often asserted: "Since different coordinate representations are just different mathematical descriptions, relevant physical entities are usually taken to correspond to coordinate-independent entities" [118, p. 1018]. On this understanding, the coordinate-dependence (i.e., non-tensorial nature) of $t^{\mu\nu}$ shows that this object is unphysical, and there can be no local notion of gravitational stress-energy. Accordingly, on this view (Par-Tt) cannot express a conservation principle relating physical quantities in GR. I dub this position 'antirealism' about non-tensorial objects, and in particular $t^{\mu\nu}$.

On the other hand, the historical Einstein did not endorse this position. Instead, Einstein maintained that the stress-energy pseudotensor could represent a physical quantity, writing "I do not see why only those quantities with the transformation properties of the components of a tensor should have physical meaning" [73, p. 167]. On this second understanding, one views $t^{\mu\nu}$ as a physical but frame-dependent quantity; I dub this 'realism' about $t^{\mu\nu}$. This is in accord with Einstein's view that the presence of a gravitational field is intimately tied to the non-vanishing of connection coefficients [71,77]:[23] since the connection coefficients are frame-dependent, it is *prima facie* plausible that gravitational stress-energy also be frame-dependent.[24] For the realist, (Par-Tt) is a legitimate but frame-dependent conservation principle.[25]

Let us reflect further on what follows if one is a realist about pseudotensorial quantities, such as $t^{\mu\nu}$. There are several questions which deserve consideration here, for example: (i) If one asserts that coordinate-dependent objects such as pseudotensors are candidates for representing real physical quantities, which coordinate system is the 'right' one for accurately so representing such quantities? (ii) Are coordinate systems supposed to be associated with observers in some way? (iii) If so, are pseudo-tensors relative quantities, like relative velocity?

One plausible line of reasoning which may be advanced in response to these questions proceeds as follows. In every coordinate system, the gravitational stress-energy pseudotensor $t^{\mu\nu}$ will take some value (possibly zero). Just as Einstein understood

[23] Though see footnote 14.

[24] This position is also advanced at [120, p. 94]; see in addition [21, p. 197] and [122, §5].

[25] For a recent attempt to make sense of (Par-Tt) as a legitimate conservation principle, see [161]. In a sense, Pitts' view is a halfway house between realism and antirealism: though he argues that pseudotensors are "physically meaningful" [161, p. 15] with "no vicious coordinate dependence" [161, p. 14], he does this by demonstrating that they can be unified into an infinite-component geometric object. For further discussion of Pitts' position, see [118, §5] and [63].

the connection coefficients $\Gamma^{\mu}{}_{\nu\sigma}$ to represent the value of the gravitational field in a given frame of reference, so too may the realist about $t^{\mu\nu}$ understand this to represent the magnitude of gravitational stress-energy in a given frame of reference. Accordingly, on this view there is no 'right' frame for accurately representing the quantity of gravitational stress-energy. Rather, on this position gravitational stress-energy is always defined with respect to a given frame of reference—so the above-suggested analogy with relative velocity is indeed apt.

4.3.2 The integral conservation law for total stress-energy

I return to whether one should be a realist about pseudotensorial quantities such as $t^{\mu\nu}$ in §4.3.3.3. In this section, however, I consider the possibility of the construction of integral conservation laws for total stress-energy in GR. Note first that (Par-Tt) can be used to construct an integral conservation law describing the interchange of stress-energy between gravitational and matter stress-energy:

$$\left(T^{\mu\nu} + t^{\mu\nu}\right)_{,\nu} = 0 \quad \Rightarrow \int_V dV \left(T^{\mu\nu} + t^{\mu\nu}\right)_{,\nu} = \int_{S=\partial V} dS (T^{\mu\nu} + t^{\mu\nu}) n_\nu = 0. \tag{4.3.1}$$

This tells us that any change in matter stress-energy (from $T^{\mu\nu}$) in a region must be balanced by an opposite change in gravitational stress-energy (from $t^{\mu\nu}$); it thereby encodes total stress-energy exchange within a volume $S \subset M$. However, as Hoefer [95, p. 194] notes, there exist here conceptual difficulties: the pseudotensorial nature of $t^{\mu\nu}$ results in (4.3.1) being ill-defined and coordinate-dependent in general. The sense in which this is so is clear: moving the part of the integral in (4.3.1) involving

$t^{\mu\nu}$ to the right hand side, we obtain

$$\int_{S=\partial V} T^{\mu\nu} n_\nu dS = -\int_{S=\partial V} t^{\mu\nu} n_\nu dS. \tag{Int-Tt}$$

Since $t^{\mu\nu}$ is coordinate-dependent in general, the same is true of the right hand side of (Int-Tt); hence the left hand side—that is, the surface integral of the matter stress-energy tensor—is also coordinate-dependent. In other words, although the integral of the sum of matter and gravitational stress-energy in (4.3.1) evaluates to zero, the (equal) magnitudes of these quantities are in general not well-defined, and hence this integral 'conservation principle' is still frame-dependent. This being said, it is important to note that there do exist physical circumstances in which one can obtain well-defined results for the left and right hand sides of (Int-Tt). One set of sufficient conditions is the following: (see [95, p. 194])

1. Integrals must be taken in the limit $r \to \infty$.
2. Asymptotic flatness of the spacetime is assumed: $g_{ab} \to \eta_{ab}$ as $r \to \infty$.
3. The coordinate system must be Lorentzian asymptotically, but can vary arbitrarily in the interior.

As Nerlich states [149, p. 159], these conditions 'impose time translation symmetry in a cryptic form'. In effect, imposing condition (2)—asymptotic flatness—allows one to treat the bulk spacetime and its content as a physical system on a Minkowski background; in this way one recovers the isometries of Minkowski space and their associated Killing vector fields, and thereby the associated conserved quantities when constructing integral conservation laws (in the $r \to \infty$ limit: condition (1)) with respect to the integral curves of these Killing vector fields (condition (3)), *à la* (4.2.13). Hence, by the work of §4.2.4, the amount of matter stress-energy in the volume V must be well-defined (with respect to the integral curves of these Killing vector fields). Then, by (4.3.1), the amount of gravitational stress-energy in the volume V must also be well-defined. In fact, this is easy to see mathematically, by contracting (Int-Tt) with the components of a Killing vector field K^a in this coordinate basis ($K_a = g_{ab}K^b$):

$$\int_{S=\partial V} t^{\mu\nu} K_\mu n_\nu dS = -\int_{S=\partial V} T^{\mu\nu} K_\mu n_\nu dS = 0. \qquad \text{(Int-TtK)}$$

Here, we have rearranged (Int-Tt) and used (4.2.13) and the symmetry of the stress-energy tensor. Hence, in this case we find that just as matter stress-energy is conserved with respect to the integral curves of the Killing vector field K^a, the same is true of gravitational stress-energy. Thus the amounts of both matter and gravitational stress-energy in the volume are well-defined, so the splitting in (4.3.1) is well-defined. This is why conditions (1)-(3) yield a well-defined (that is, frame-independent) conservation principle for total stress-energy.

When conditions (1)-(3) hold, (4.3.1) can be applied in order to calculate (for example) stress-energy loss by a system due to gravitational wave transportation; such calculations agree with observations on binary star/pulsar pairs [95, p. 194]. Hence, in such cases there appears to exist a well-defined quantity (with respect to a class of frames) which balances any change of matter stress-energy of the system, in exactly the same manner as in SR. Regardless of whether one is a realist or antirealist about $t^{\mu\nu}$ in the sense above, this new quantity is *prima facie* a candidate for a well-defined notion of gravitational stress-energy in GR.[26]

Finally, with the above in hand, we are in a position to understand why some

[26] Localisability is a stronger condition than satisfying an integral conservation principle, because it is possible to have the latter in the absence of the former, as is the case in (4.3.1) when satisfying (1)-(3).

authors (for example [7, 95, 118]) have claimed that only integral conservations laws should be considered conservation principles 'properly speaking', *contra* §4.2.2.3. It is likely that such claims are made *post hoc*, in light of the frame-dependence of expressions such as (Par-Tt). However, there exist at least two issues with this view. First, such a position is highly revisionary: it is more in line with physical practice to state that both differential and integral conservation laws are *a priori* legitimate; it is only if these contain non-tensorial quantities and one is an antirealist about these quantities that one can claim that such conservation laws are not genuine. Second, this position faces the obvious objection that some integral conservation laws such as (4.3.1) are also generically ill-defined.[27]

4.3.3 Gravitational stress-energy

The logic of the previous two subsections was as follows: Though (Par-Tt) has the form of a differential conservation law (Par-T), one of its relata is a frame-dependent quantity. Whether one views (Par-Tt) as a legitimate conservation law (that is, as a conservation law relating physical entities) will therefore depend upon whether one thinks that frame-dependent mathematical objects such as $t^{\mu\nu}$ can represent physical quantities. Whatever one's take on this though, it is also true that in some physical circumstances, an integral version of (Par-Tt)—that is, (Int-TtK)—appears to relate frame-independent physical quantities, one of which corresponds to a notion of gravitational stress-energy.[28] The question to be pursued now is whether such a quantity does indeed represent gravitational stress-energy as a physical magnitude.

[27] This may be evaded straightforwardly if one claims that being an integral conservation law is a necessary but not sufficient condition to be a conservation principle 'properly speaking'.

[28] Recall from §4.2.4 that, while conserved quantities such as energy strictly only exist in frames related by the appropriate coordinate transformations, by projecting onto the Killing vector fields associated with such transformations, one can obtain results which hold in any frame. It is this which I mean by a 'frame-independent' notion of gravitational stress-energy—even though strictly such stress-energy (*sans* projection) is conserved only in a restricted class of frames.

In this subsection, I evaluate two positions in the literature on this point. According to the former (advocated by Hoefer [95]) there exists no genuine gravitational stress-energy in GR in any sense. According to the latter (advocated by Lam [118]), there does exist frame-independent gravitational stress-energy in GR, in the sense that the conservation law (Int-TtK) holds in some models of the theory. Though both Hoefer and Lam take an antirealist attitude towards $t^{\mu\nu}$ (in the sense of §4.2.5), the position of Lam is compatible with a realist understanding of $t^{\mu\nu}$.

4.3.3.1 Against weak gravitational stress-energy?

Based upon the results of the previous section, Hoefer argues that in GR: (a) the stress-energy of the gravitational field is ill-defined both locally (that is, at a point) and globally (that is, in the sense of an integral conservation law); and (b) there is no general principle of total stress-energy conservation in GR. He claims: "[W]e should abandon this effort to gloss over the facts. Let the textbooks admit openly that gravitational field stress-energy is not well-defined or fundamental, and that neither it nor ordinary stress-energy is conserved" [95, p. 195].

Hoefer adopts an antirealist line regarding pseudo-tensorial quantities, and thereby rejects both (Par-Tt) and (4.3.1) as conservation principles. (While Hoefer also argues that equations such as (Par-Tt) should be rejected as conservation principles on the grounds that they are not integral conservation laws, we have seen above that such reasoning is misguided.) With this in mind, the most interesting of Hoefer's claims is the assertion that genuine gravitational stress-energy does not exist in GR even when a frame-independent quantity playing this functional role exists in the theory, as with (Int-TtK). Hoefer argues that the stringent limitations on the applicability of (Int-TtK) imposed by (1)-(3) make this no genuine conservation principle in these circumstances either, for two reasons:

(i) The actual world is not asymptotically Minkowski, so "[(Int-TtK)] does not apply to gravity in the actual world" [95, p. 194].

(ii) Holding (Int-TtK) as an important physical result "goes against the most important and philosophically progressive approach to spacetime physics: that of downplaying coordinate-dependent notions and effects, and stressing the intrinsic, covariant and coordinate-independent as what is important" [95, pp. 194-195].

What should one make of these two claims? Beginning with (i), one might object to this on various grounds. First, the first statement of (i) is undeniable in the sense that the entire universe is not asymptotically Minkowski. Nevertheless, this does not preclude us from applying (Int-TtK) when modelling certain physical systems in the actual world (for example binary star systems). Hence, it appears that the second claim of (i) does not follow from the first, at least when physical systems within the world are considered in isolation.[29] A second, related reason to object to (i) is the following: every theory of physics is an idealisation and does not 'apply to the actual world' in this strong sense. So, Hoefer's objection levied at (Int-TtK) seems at the same time to apply to an unacceptably broad class of physical laws and theories.

In addition, one might object to (i) on the grounds that our concern is not with the specific DPM of GR (see §4.2.1) which is taken to model the (cosmology of) the actual world, but rather with the entire space of DPMs of GR. If a frame-independent physical quantity corresponding to gravitational stress-energy can be defined in a certain subclass of those DPMs, then that is sufficient to conclude that a frame-independent

[29]In other words, the notion of gravitational stress-energy may still be applicable to (subsystems of) the actual world at an approximate, functional level—see §4.3.3.3 below.

notion of gravitational stress-energy does exist in GR. On this way of understanding the dialectic, consideration of the actual world is broadly irrelevant to the question of whether a frame-independent notion of gravitational stress-energy exists in GR.

Once this point is recognised, one is also capable of responding to any objection to the statement that a frame-independent notion of gravitational stress-energy exists in GR on the grounds that the models of the theory in which this notion may be defined are 'rare', or 'unstable'—in the sense that perturbing the model slightly yields a new model of GR in which such a notion of frame-independent gravitational stress-energy cannot be applied.[30] The nature of this response is straightforward: such issues are (once again) irrelevant to the question of whether a frame-independent notion of gravitational stress-energy can be defined in GR simpliciter.

On (ii), this claim is again objectionable. First, it is clear that the statement is a mixture of an appeal to a majority view (if indeed this is a majority view, as Hoefer asserts) and a re-assertion of the antirealist position presented in §4.3.1, with no concrete argument presented for this position. Indeed, even if Hoefer can argue for this antirealist position, it is not clear it applies in the case under consideration, that is (Int-TtK). This is because here we have a frame-independent notion of gravitational stress-energy, which seems to match the desiderata laid out in (ii) anyway!

Perhaps Hoefer has in mind the following worry: although in such cases we appear to have a frame-independent notion of gravitational stress-energy, this is only after projecting onto Killing vector fields. In fact, as we saw in §4.3.2, stress-energy in GR (*sans* such projection) is only a well-defined quantity in a restricted class of frames. While this is true, commitment to the view that we do not have total stress-energy conservation law in GR due to the fact that the relevant conservation princi-

[30]For example, in a model of GR in which conditions (1)-(3) are satisfied, it suffices to introduce a small perturbation in the metric field such that it is not asymptotically Minkowski at infinity for the above notion of frame-independent gravitational stress-energy to no longer be applicable.

ples do not hold in every frame leads to potentially undesirable consequences. Most notably, such a claim would also commit one to the statement that there exists no genuine stress-energy conservation law in SR—a theory in which the conservation of total stress-energy typically is taken to be uncontroversial. While Hoefer is free to adopt such a position, he does not appear to do so (see, for example, [95, p. 189]).

A further worry regarding such general antirealism about gravitational stress-energy in GR is the following. As Baker [7, p. 1305] notes, the advocate of a Hoefer-type view is apparently committed to the denial of the claim that gravitational waves and other forms of purely gravitational radiation are energetic. For example, for the case of a binary star system where conditions (1)-(3) hold and (Int-TtK) is well-defined, textbook accounts state that the matter stress-energy of the system decreases as some stress-energy is carried away in gravitational radiation (see, for example, [91, 141, 196]). For Hoefer, such a story cannot be told. Instead, he will have to assert that the matter stress-energy of the system just decreases, and stress-energy is not conserved. Though it is likely that Hoefer will bite the bullet on this point, it is certainly a revisionary view.

4.3.3.2 Gravitational stress-energy relative to background structure

The above in mind, it does not appear that strong reasons have been given to support the view that no genuine gravitational stress-energy exists in GR. Let us now lay out an alternative perspective on gravitational stress-energy, presented by Lam, who makes the weaker claim that "the very notion of energy—gravitational or not—is well defined in the theory only with respect to some background structure" [118, p. 1012]. What is meant by 'background structure' here? Suppose that for a given $\langle M, g_{ab}\rangle$ there exists a Killing vector field K^a. Then there exists an isometry $\phi_* g_{ab} = g_{ab}$ generated by K^a, where ϕ is a diffeomorphism along the integral curves of K^a. In this sense, the Killing vector field is associated with spacetime structure 'stationary' under ϕ, and can thereby be taken to indicate non-dynamical 'background structure' with respect to which integral stress-energy conservation can be demonstrated. By contrast, a fully dynamical metric will in general lack the above stationarity and so preclude the existence of integral conservation laws.

With the above characterisation of 'background structure' in hand, Lam's position can be presented as follows: (a) conserved quantities such as stress-energy are only well-defined in the presence of a Killing vector field (the 'background structure'); (b) conditions (1)-(3) provide a specific situation in which Killing vector fields can be constructed in GR, with associated integral conservation laws; (c) in such cases, the stress-energy (including gravitational stress-energy) associated with those Killing vector fields, constructed via the integral conservation law (Int-TtK), is a frame-independent quantity and therefore (Lam claims) genuine. In light of this, Lam concludes "in the cases in which total energy-momentum is well defined (and conserved), it is a global notion" [118, pp. 1022-1023]).

The advocate of this view will therefore maintain that there appear to be contexts in which sufficient 'background structure' exists that conditions such as (1)-(3) hold for the system under consideration; here a frame-independent quantity representing gravitational stress-energy is well-defined, and in such contexts (including real-world contexts, such as binary star systems[31]) it does make sense to speak of the gravitational stress-energy of the system, and of (4.3.1) as being a legitimate conservation law. Though I concur with Lam on this point, I diverge in the interpretation of pseudotensorial quantities such as $t^{\mu\nu}$—see below.

[31] At least at an approximate level—see footnote 29.

4.3.3.3 Functionalism about gravitational stress-energy

The above two positions in mind, one must ask two questions: (a) is it correct to call the quantity appearing in (Int-TtK) associated with $t^{\mu\nu}$ 'gravitational stress-energy', and (b) does such 'gravitational stress-energy' really exist in GR?[32] Begin with (a). As Lam notes, "energy and mass might not be fundamental properties of the world, in the sense that they make sense only in some particular (but very useful) settings; this does not lessen the fact that the notions of energy and mass constitute extremely powerful tools for many concrete and practical cases" [118, p. 1023].[33] This point is important: gravitational stress-energy *à la* (Int-TtK) is not a fundamental concept in GR, insofar as it is only applicable in a limited range of DPMs of the theory—namely, those in which certain Killing vector fields may be defined. (That is, the term 'gravitational energy' is associated with structures—namely terms such as that on the left hand side of (Int-TtK)—which are not used to construct the space of DPMs of GR, but rather which are only well-defined in a certain subset of those DPMs.) Nevertheless, in such instances it is extremely useful to make use of this term, within that subclass of DPMs. Hence, at a practical level, it is legitimate to call such a quantity gravitational stress-energy.

Clearly though, this does not settle the putative ontological issue (b), concerning whether gravitational stress-energy 'really' exists in GR. In my view, it is plausible to maintain that in situations such as those in which (Int-TtK) holds, there exists a quantity in GR which fulfils the functional role of gravitational stress-energy. The reasons for this are the following. First, for a structure in a certain model of a theory to play the 'functional role of gravitational stress-energy', it must (i) fulfil a function analo-

[32]Note that (b) is not the same as asking whether gravitational stress-energy really exists in the (possibly general relativistic) actual world, for the reasons delineated in §4.3.3.1.

[33]Here, Lam is referring to both matter and gravitational (stress-)energy. Note also that Lam says 'properties of the world', rather than 'properties in GR'. For my purposes, it is legitimate (and preferable) to read him as making the latter, more general statement (see footnote 32).

gous to that of gravitational energy as traditionally conceived—namely, as a quantity (gravitational potential energy) in Newtonian gravitation, which balances the matter (in Newtonian mechanics: kinetic) energy of the system in question such that their sum is conserved; and (ii) bear some relation to the 'gravitational' degrees of freedom in the theory in question. Second, arguably terms such as that on the left hand side of (Int-TtK) do satisfy (i) and (ii)—the former holds in virtue of a comparison of the form of (Int-TtK) with Newtonian energy conservation equations; the latter in virtue of the connections between $t_{\mu\nu}$ and, for example, the connection components—themselves (at least, on views such as those indicated by Einstein mentioned above) understood to be associated with 'gravitational' degrees of freedom, as elaborated in §§4.2.3.3 and 4.3.1.[34] A functionalist may, therefore, speak of the existence of gravitational stress-energy in such situations.[35] On the assumption that (i) and (ii) are satisfied, the alternative to functionalism is to say that 'the structure of certain DPMs of GR is such that it appears that there exists gravitational stress-energy in those models, but really there is no such stress-energy there'; the payoff to be gained from making such a claim is unclear.

Still, doubts may linger. In particular, one might argue as follows: 'Surely there is a much more plausible alternative that disputes that gravitational energy 'really' exists, which says that we can describe everything that is going on in terms of solutions to (4.2.1) (and its consequences, including (Cov-T)), without any need to help ourselves to talk involving $t_{\mu\nu}$.' With this point, I am in broad agreement: one could indeed explain all general relativistic phenomena, in any model of the theory, simply using the apparatus used to pick out the DPMs of the theory. Nevertheless, I respond that

[34]Though one should recall the caveats of §4.2.3.3 regarding the question of which object in GR should be associated with the 'gravitational field', strictly speaking.

[35]Such a line accords with general functionalist attitudes in science. Wallace [215, p. 58] summarises this as follows: "Science is interested with interesting structural properties of physical systems, and does not hesitate at all in studying those properties just because they are instantiated 'in the wrong way'."

there may be other well-defined structures applicable only in certain models of the theory, which play certain functional roles. As I see it, there is nothing illegitimate in regarding such structures as also existing in (the worlds represented by) those models of the theory—and, moreover, doing so may open up more perspicuous avenues for the explanation of phenomena within those models (recall from §4.3.3.2 the case of gravitational radiation from binary star systems). Thus, while I concur with the above argument, in my view this does not constitute an objection to a notion of functional gravitational energy in GR, since advocates of such a concept simply have more explanatory apparatus available to them.

Thus, a functionalist may assert that gravitational stress-energy in the sense of (Int-TtK) does exist in GR, but only in a restricted class of situations; this aligns with the case of SR. Since we have already found Hoefer's objections to the existence of gravitational stress-energy in GR wanting, and since such functionalist principles are practical and simple, I conclude that in such situations (corresponding to certain models of GR) it is best to state that a quantity which plays the functional role of gravitational stress-energy does exist in the theory, and hence should be labelled as genuine. I conclude that frame-independent gravitational stress-energy does exist in GR, in the limited sense above.

Whether one also maintains that a frame-dependent notion of gravitational stress-energy exists in GR will, for the reasons discussed, depend upon whether one is a realist or antirealist about pseudotensorial quantities, in particular $t^{\mu\nu}$. Note, however, that realism about such quantities is arguably compatible with the above functionalist principles: in each given frame of reference, one may define a quantity, represented by $t^{\mu\nu}$, such that total stress-energy is conserved. Again, this plays the functional role of gravitational stress-energy in that frame, for (i) and (ii) as delineated above are both satisfied. Thus, it is also the case that speaking of frame-dependent gravitational stress-energy may be justified on functionalist grounds.

On this latter point, there exist connections with other, ongoing debates in the foundations of spacetime theories—in particular, over the primacy of coordinate-dependent versus -independent explanations. According to advocates of the latter, such as Friedman [81] and Maudlin [137], explanations of physical phenomena within models of a given spacetime theory should proceed by appeal only to coordinate-independent structures. By contrast, according to advocates of the latter, such as Brown [24] and Wallace [217], presentations of spacetime theories need not proceed in a coordinate-independent manner; rather, spacetime theories may be defined in terms of equations written in a coordinate basis and their transformation properties (this is what Brown [24, p. 9] and Wallace [217, p. 5] refer to as the 'Kleinian conception of geometry'[36]), and explanations may be given by appeal to those laws, written in a coordinate basis. On the former programme, in the context of, for example, SR, explanations of phenomena proceeding by appeal to coordinate-dependent effects (for example, the twin paradox differential in terms of the relativity of simultaneity, assuming some clock synchrony convention) are to be rejected, as the associated physical effects not considered 'real'; on the latter approach, there is nothing wrong with issuing such explanations, and with viewing such effects as physical. Those who buy into the latter programme may view the notion of gravitational energy in question as frame-dependent, but no less real for all that.

[36] I set aside here the question of the extent to which this 'Kleinian conception' is faithful to the definition of geometry in Klein's *Erlangen* program [109].

4.4 Gravitational Stress-Energy and Spacetime Ontology

4.4.1 Relationism and gravitational stress-energy

In the above, I endorsed a position according to which gravitational stress-energy does exist in GR, in at least (i) a weak sense applicable in a certain class of frames of a certain class of models of the theory, and arguably also (ii) a strong sense, applicable in all frames of all models of the theory. Nevertheless, there exists a residual worry, related to the debate between substantivalism and relationism about spacetime ontology. In this chapter, I take substantivalists to claim that spacetime exists as an entity in its own right, and relationists to deny this—that is, to claim that all talk of spacetime is reducible to talk of (relational properties of) matter fields. This in mind, the worry regarding gravitational stress-energy runs as follows: if one is a relationist, how can one maintain that there indeed exists gravitational stress-energy in the world? The thought that relationism implies the nonexistence of gravitational stress-energy (or, equivalently, that genuine gravitational stress-energy implies substantivalism) is intuitive. In this section, however, I argue that it is not correct.

There are two different pictures of the substantivalism/relationism debate which are relevant here. Let $\langle M, g_{ab}, \Phi \rangle$ be a model of GR. Then, according to 'manifold substantivalism', spacetime is identified with M; to be a relationist is to maintain that manifold points do not have an ontological status independent of the fields defined upon them. On the other hand, according to 'metric substantivalism', a spacetime is identified with $\langle M, g_{ab} \rangle$—that is, with both the manifold and the metric field. To be a relationist is then to maintain that neither manifold points nor the metric field have a distinct existence over and above the matter fields Φ. Clearly, it is harder to be a relationist in the second sense than the first.[37]

For the purposes of this chapter, it is more relevant to consider relationism in the second sense above, since we are concerned with the ontological status of quantities associated with g_{ab}. In this case, the question becomes: how can those who do not believe that the metric field is fundamental (insofar as they think it reducible to properties of matter fields) maintain the existence of genuine gravitational stress-energy? On the face of it, such a claim is implausible, and indeed many (for example, [7, 95]) maintain that the answer is a simple negative: they cannot.

I answer to the contrary. My positive story runs as follows. Whether one is a relationist or a substantivalist, it is a fact that there are some situations in which results such as (Int-TtK) are well-defined, and there appears to be a well-defined quantity in the theory which plays the functional role of frame-independent gravitational stress-energy. Of course, the account given by the relationist of this quantity will differ from that given by the substantivalist: the relationist will assert that this quantity is associated (in some way to be made precise) with the metric field, which is in turn a codification of properties of the fields; the substantivalist will appeal to the metric field simpliciter. In either case though, this quantity exists in the theory: the two sides are not debating its existence, but rather the fields to which it is ultimately attributable.

[37]Two points are in order here. First, work such as [121] demonstrates that the stress-energy tensor T^{ab} of the matter fields Φ of GR in fact presupposes metric structure in its definition. It is perhaps then misguided to attempt to reduce g_{ab} to T^{ab}. That said, one should of course recall that one should not conflate matter fields Φ with their associated stress-energy tensor T^{ab}. Thus, even if it *is* misguided to attempt to reduce g_{ab} to T^{ab}, perhaps it is still possible to reduce g_{ab} to Φ.

The second point is related: suppose one does seek to reduce g_{ab} to Φ. To do so is to endorse a version of 'Mach's principle' [122, p. 455]. There exist problems with attempting to achieve this in GR—for example, one faces the problem that *a priori* the metric field has ten independent components, whereas the electromagnetic field tensor (for example) has only six—so there is a question of how the former can be reduced to the latter. Another problem in this vicinity lies in the existence of vacuum solutions in GR (cf. Ch. 1). As a result, it is questionable whether the form of relationism considered here is ultimately defensible in GR—so the work of this section is best viewed as (a) a response to [7,95], where it is claimed that such views are incompatible with the existence of gravitational energy; and (b) an exploration of the consequences of the functionalism about gravitational stress-energy developed above.

4.4.2 The cosmological constant

Finally, it is worth commenting on Baker's claims regarding the bearing of the possibility of a non-zero cosmological constant Λ on both the substantivalism/relationism debate, and the existence of gravitational stress-energy [7, §4]. Including a cosmological constant term, (4.2.1) reads

$$G_{ab} + \Lambda g_{ab} = 8\pi T_{ab} \tag{4.4.1}$$

Let us focus on the following two claims made by Baker:

1. $\Lambda \neq 0$ commits us to (manifold-plus-metric) substantivalism, because this quantity is associated with spacetime yet cannot be reduced to relations amongst the matter fields [7, §4.1].
2. $\Lambda \neq 0$ results in a non-zero vacuum energy density, for which a relationist cannot account [7, §4.2].

(1) questions whether relationism is compatible with a non-zero cosmological constant. (2) assumes that such is the case, but claims that relationism nevertheless cannot account for the vacuum energy density arising from non-zero Λ. Let us first consider (1). In fact, there are several reasons to be suspicious of this claim. First, Baker justifies that Λ is 'associated with spacetime' as follows: [7, p. 1301]

> [Λ's] role in the field equations [(4.4.1)] is to influence, by itself or in combination with other terms, the metric structure of spacetime, and thereby to affect the physical behaviour of matter. This is exactly the sort of influ-

> ence that accounts for gravitational forces in GR, the only difference being that Λ does not depend on matter as its source.

This claim is suspect, because Baker has not ruled out the possibility that the Λ term in (4.4.1) can appear on the right hand side of this equation, with Λ being treated as another matter field; such an interpretation of (4.4.1) is also *prima facie* possible, yet the above assertion does nothing to rule it out. Therefore, Baker needs to do more to make convincing any claim that Λ is 'associated with spacetime'. Indeed, there is an ongoing debate in cosmology over whether to consider the cosmological constant term in (4.4.1) as being attributable to spacetime (that is, as sitting on the left hand side), or as another type of matter-like field (that is, as sitting on the right hand side); Baker cannot simply presuppose an answer to this question.

Second, the claim that Λ cannot be reduced to relations among the matter fields is also too quick. On this, Baker states: [7, p. 1306]

> I do not doubt that a persistent relationist could describe Λ's effects as mere relational properties, but the price will be high. Considering [the] example of distant objects moving apart under the influence of Λ, the relationist would have to posit a brute fact that material objects possess a tendency to accelerate away from one another at a rate proportional only to the distance between them.

In fact, this seems to be roughly in line with the relationist-type approach to the metric field in relativity theory in the context of SR outlined in [24]. More generally, Baker cannot infer from the fact that such a programme appears to be undesirable to him that it cannot be done (and indeed, he openly admits that it can be done), or even that such a task might not be desirable or acceptable in some relationist research programmes.

On (2), Baker claims that a relationist cannot account for the vacuum energy density $\rho_\Lambda = \Lambda/8\pi G$ implied by a non-zero Λ:[38] "I can see no easy way for the relationist to explain the energy density of empty space" [7, p. 1310]. In fact though, this is not so: the relationist can account for the energy density of empty spacetime. To see this, suppose that the cosmological constant can be reduced to properties of matter fields. Then, as with gravitational stress-energy in the previous subsection, the equations of the theory will still state that there exists a quantity which plays the functional role of a vacuum energy density. Once again, the only difference will be the story that is told to account for this quantity. While the substantivalist will appeal directly to g_{ab} and Λ, the relationist will resort to an elliptic story about how this quantity arises from properties amongst the matter fields themselves. But on either account, a functional vacuum energy density exists according to the theory.

4.5 Conclusion

In this chapter, I have reconsidered the existence of gravitational stress-energy in GR; adopting a functionalist attitude to the definition of physical quantities, I have argued that gravitational stress-energy can be considered to exist in GR, in both (i) a weak sense applicable in a certain class of frames of a certain class of models of the theory (namely, models which instantiate certain symmetries, and therefore which possess Killing vector fields), and (ii) arguably also in a stronger sense (as represented by the gravitational stress-energy pseudotensor in a given frame), applicable in all frames of

[38] As Baker states, this does not mean that Λ *arises* from the non-zero ρ_Λ; rather, its role in the field equations is equivalent to an energy density of empty space [7, p. 1309].

all models of the theory. This latter approach runs against contemporary orthodoxy, but is in line with the thinking of the historical Einstein [73, p. 167].

In addition, I have adopted a revisionary line regarding whether gravitational stress-energy is compatible with relationism, arguing that regardless of whether one thinks that the metric field g_{ab} is reducible to properties of matter fields, gravitational stress-energy still exists in the theory, if one again embraces functionalism about physical quantities. Accordingly, one's position on the ontology of spacetime does not affect one's commitment to gravitational stress-energy in GR; this point also applies to the claim that a non-zero vacuum energy density is incompatible with relationism.

Part III

Symmetries and Dualities

Chapter 5

Duality and Ontology

A 'duality' is a formal mapping between the spaces of solutions of two empirically equivalent theories. In recent times, dualities have been found to be pervasive in string theory and quantum field theory. Naïvely interpreted, duality-related theories appear to make very different ontological claims about the world—differing in e.g. spacetime structure, fundamental ontology, and mereological structure. In light of this, duality-related theories raise questions familiar from discussions of underdetermination in the philosophy of science: in the presence of dual theories, what is one to say about the ontology of the world? In this chapter, I undertake a comprehensive survey of the landscape of possible ontological interpretations of duality-related theories.

5.1 Introduction

Contemporary physics is built upon two *prima facie* mutually incompatible frameworks: the theory of general relativity on the one hand, and the standard model of particle physics—a certain quantum field theory—on the other. Although these two frameworks are strikingly effective at describing the actual world in their relevant domains (*viz.*, astrophysical and cosmological scales for general relativity, and atomic and subatomic scales for the standard model), they rest upon very different assumptions. For example, one central feature of general relativity is that spacetime is rendered dynamical (that is, it is not a fixed background, but rather 'curves' in response to its matter content); by contrast, spacetime remains a fixed background in the standard model of particle physics.

This notwithstanding, various fields of physics—such as the study of the early universe, or of black holes—lie at the intersection of the domains of these two theories, and thereby call for a *quantum theory of gravity*.[1] Constructing such a theory capable of overcoming the tensions between general relativity and the standard model is an ongoing matter of profound difficulty; at present, there exist several candidate options which remain the subject of active research.[2]

According to the naïve ontological picture presented by *string theory*—arguably the most popular extant research programme in quantum gravity—reality is constituted by one-dimensional strings, as well as by other higher-dimensional entities called 'branes'. Moreover, reality is not made up of four spacetime dimensions (three spatial and one temporal), but rather of ten, or eleven. But string theory embodies another intriguing notion that should be of profound interest to philosophers and metaphysicians—the notion of *duality*.

[1]Strictly, one might distinguish different senses of a 'quantum theory of gravity'—this could be e.g. (i) a quantised version of a theory which describes gravity, such as a quantised version of general relativity; or (ii) a theory which unifies general relativity and the quantum-mechanical standard model of particle physics; or perhaps (iii) something else. (Sometimes, theories of type (ii) are called 'theories of everything', since they encompass all four fundamental interactions.) It is the notion of a theory of quantum gravity of type (ii) which is my concern here. (For further discussion on these matters, see [101,236].)

[2]For a philosophical overview of such options, see [103].

Associate with every theory a class of 'models', equivalently 'solutions'. As van Fraassen puts it, a model is "Any structure which satisfies the axioms of a theory" [210, p. 43]. (We have already seen much of this framework with our discussion of KPMs and DPMs in the preceding chapters of this thesis.) In turn, take it that two solutions are 'empirically equivalent' just in case they agree on all 'physically observable data', i.e. on empirical substructures, in the sense of van Fraassen [210, p. 64]. Then, a duality is a mapping between the spaces of solutions of two theories, such that models related by that map are empirically equivalent.[3]

There exists not just one string theory, but rather five, related by an intricate web of dualities.[4] That is to say, each model of a given string theory possesses (via the duality map under consideration) a dual model, which *prima facie* makes very different ontological claims about the world, while nevertheless being empirically equivalent. The four best-known examples of dualities arising in string theory are 'T-duality', 'mirror symmetry', 'S-duality', and the 'AdS/CFT correspondence'. In the case of T-duality (already discussed in the context of the bosonic string in Ch. 2), type IIA superstring theory (one of the five superstring theories) on a product manifold $M \times S^1$ with radius of the periodic dimension R is found to be dual to type IIB superstring theory (another of the five superstring theories) on the product manifold $M \times S^1$ with radius of the periodic dimension proportional to $1/R$ [14, ch. 6]. Mirror symmetry is a generalisation of T-duality to the case of topologically inequivalent man-

[3]Such is the definition of dualities presented in e.g. [131, 178], which will suffice for my purposes in this chapter and the next. Note that one might augment the criterion of 'empirical equivalence' by requiring that all quantities regarded as being physically meaningful (whether observable or not) be preserved under the duality map. I agree that the most striking examples of dualities—including examples of dualities from string theory—satisfy this stronger definition. However, in this thesis I choose to work with the weaker notion of a duality proceeding in terms of empirical equivalence alone, for this will suffice to make all necessary points regarding the interpretation and ontology of duality-related theories. For more detailed and comprehensive approaches to the definition of dualities, see e.g. [48, 51]. Note also that it need not be the case that the duality map is one-one—it might instead be that a class of solutions of the first theory is mapped to a single solution of the second theory under the duality. This will be of relevance below.

[4]Since these string theories incorporate supersymmetry, they are sometimes known as 'superstring theories'.

ifolds. (For a philosophical introduction to mirror symmetry, see [186].) S-duality relates solutions of one superstring theory with string coupling constant g_s to solutions of another superstring theory with string coupling constant $1/g_s$; it is thus a so-called 'strong/weak' duality. For example, strongly/weakly coupled type I superstring theory is dual under S-duality to weakly/strongly coupled $SO(32)$ heterotic string theory [14, §8.2]. Finally, in the AdS/CFT correspondence, a string theory in so-called 'AdS spacetime' (the 'bulk theory') is dual to a conformal field theory (CFT) in a lower number of spacetime dimensions (the 'boundary theory') [14, ch. 12].[5] In this chapter, the relevant aspects of these dualities will be introduced when needed in the course of the dialectic.

At first blush, dualities instantiate the *underdetermination of theory by empirical evidence* familiar from the philosophy of science. In the case of dualities, however, this underdetermination is peculiar, as the empirical equivalence of the solutions under consideration was often not expected *ab initio*, but rather came as a profound surprise, in light of their apparently diverging ontological pictures.[6] In this chapter, I undertake a comprehensive survey of the terrain of possible interpretative options for ascertaining the ontology of duality-related models of physical theories in such a way as to resolve any threat of underdetermination, introducing several novel observations and options along the way.[7]

[5]The AdS/CFT correspondence was originally introduced by Maldacena in [129].

[6]This is, at least, the case for the superstring dualities—though of course, given one theory, one *could* construct retrospectively a class of dual theories (here, the Poincaré disc model comes to mind—cf. [163]).

[7]For other philosophically-oriented introductions to dualities—including all the theories and their respective dualities mentioned in this chapter—see e.g. [166, 184]. For an introduction to the philosophy of quantum gravity more generally, see [132].

5.2 Review

In the recent literature, one finds the above-mentioned claim that string-theoretic dualities present a case of underdetermination of theory by evidence—that is, a situation in which there exist multiple theories, each of which (*prima facie*) makes different ontological claims about the world, yet which are all adequate to exactly the same stock of (possible) empirical data. Such underdetermination is typically understood to be problematic for the scientific realist, for how can one plausibly maintain that one's preferred theory is *true*, if a range of other theories are also consistent with the data? In e.g. [131,176], authors began to compile a taxonomy of interpretative options available to the realist, in order to 'break' the putative underdetermination arising in the case of dualities. The (allegedly) most plausible options in this regard were claimed to be the following:

- (*Discrimination*.) Privilege the ontological claims of just one of the two dual theories. That is, consider two dual theories, $\mathcal{T}_1$ and $\mathcal{T}_2$, with (respectively) solutions $\mathcal{M}_1$ and $\mathcal{M}_2$ related by the duality map. Naïvely interpreted,[8] $\mathcal{M}_1$ and $\mathcal{M}_2$ represent two *distinct* worlds, respectively W_1 and W_2 (hence a case of underdetermination). However, according to this *discriminatory* strategy, only one of W_1 and W_2 is a legitimate description of the actual world.[9] Though coherent,

[8]What do I mean, when I speak of the 'ontological claims' of (solutions of) a theory, or of the 'naïve interpretation' of (solutions of) that theory? In this thesis, I take the 'ontological claims' of a theory to be given by its 'naïve interpretation'; in turn, I understand this to be an interpretation of the theory in question (*a fortiori* its solutions) such that the worlds represented by the solutions of that theory are 'isomorphic' to those solutions. (Here, I set aside legitimate concerns that speaking of isomorphism between mathematical structures and worlds constitutes a category error—cf. [212, ch. 1].) More technically, my focus is upon *internal* interpretations, in the sense of [48,55].

[9]Throughout this chapter, by the 'actual world', I mean a hypothetical world in which the empirical data consistent with two dual solutions are observed. Though this use is somewhat non-standard, it will simplify the discussion. It is also worth clarifying what I mean by 'legitimate description'. Suppose that one observes certain empirical data, and one has to hand a certain range of mathematical models consistent with that data. In spite of all such models being consistent with the data, it may be that one discounts certain models, for certain super-empirical, philosophical reasons (more on this

this approach faces an obvious problem: principled reasons for privileging the ontological claims of just one of the two dual solutions appear (in general) to be lacking (cf. [176, 204]).

- (*Common core*.) 'Break' the underdetermination by interpreting only the 'common core' of the solutions related by the duality map as representing physical states of affairs. In more detail, consider again two dual theories, $\mathcal{T}_1$ and $\mathcal{T}_2$, with (respectively) solutions $\mathcal{M}_1$ and $\mathcal{M}_2$ related by the duality map. On this position, the 'naïve' interpretation of $\mathcal{M}_1$ and $\mathcal{M}_2$, according to which these solutions represent distinct worlds W_1 and W_2, is not correct. Rather, we should identify the mathematical structure *common* to those solutions, and interpret $\mathcal{M}_1$ and $\mathcal{M}_2$ in terms of only that common structure—call it $\mathcal{M}_c$. In so doing, the underdetermination is (apparently) broken, for in so interpreting $\mathcal{M}_1$ and $\mathcal{M}_2$, these solutions may be regarded as representing the *same* world—call it W_c—the ontology of which is taken to be represented by $\mathcal{M}_c$.

In this chapter, my concerns are twofold: (1) I contend that both the discriminatory and common core approaches are more subtle than has hitherto been appreciated—and in fact, both approaches are consistent with a number of *distinct*, more fine-grained views, only some of which overcome the putative underdetermination in the case of dualities. (2) I maintain that there exist (at least) two *further* approaches for addressing the underdetermination which arises in the case of dualities—these I call 'nihilism' and 'pluralism'. Roughly speaking, nihilism is the view that *no* solutions of dual theories constitute legitimate descriptions of the actual world;[10] pluralism is the view that *all* dual solutions may be taken to represent *the same* actual world—but

below). Such models are, then, taken to *not* constitute legitimate candidates for representing the actual world. The complement of this set of available models is the set of legitimate candidates for representing the actual world.

[10]One might wonder how nihilism can be compatible with scientific realism; we shall see in §5.7 multiple senses in which this could be the case.

not because such a world is represented by the common core of those solutions, but rather because the structure of all dual solutions may be instantiated *simultaneously*.[11]

In the remainder of this chapter, I undertake the following tasks. In §5.3, I propose an expanded taxonomy of options for the interpretation of dualities. In §5.4, I present and set aside various 'antirealist' and 'structuralist' approaches to dualities—for my concern in this chapter is to interpret dual theories realistically; that is, to get a handle on what dual theories tell us about *what the world is really like*. In §§5.5-5.8, I discuss each of the above-mentioned realist interpretative options in turn, and assess whether they succeed in resolving the putative underdetermination arising from dualities.

5.3 Taxonomy

What I call above 'discrimination' and the 'common core' approach form just two of a substantially broader range of interpretative options *vis-à-vis* dualities. To make this explicit, consider figure 5.1. Here, $\mathcal{M}_1, \ldots, \mathcal{M}_5$ represent (respectively) five solutions of five theories $\mathcal{T}_1, \ldots, \mathcal{T}_5$, which are dual to one another.[12] The solution $\mathcal{M}_c$ to the right of the $\mathcal{M}_1, \ldots, \mathcal{M}_5$ consists in the common mathematical structure of each of the five dual solutions.[13,14] Beneath each of the $\mathcal{M}_1, \ldots, \mathcal{M}_5$ are sets of possible worlds (1)-(6), to which the solutions are interpreted as corresponding.[15] If such a world is to

[11]The idea of pluralism is owed to Baptiste Le Bihan.

[12]I here include five solutions with an eye to the five superstring theories. However, any number of dual solutions greater than or equal to two would suffice for my purposes.

[13]The rightward arrow indicates that $\mathcal{M}_c$ is typically constructed once the $\mathcal{M}_1, \ldots, \mathcal{M}_5$ are given.

[14]What exactly does this 'common mathematical structure' consist in? This is a question worthy of considered attention; thankfully, authors in the philosophy of dualities have already done much to clarify these matters. See in particular [51, §2], in which this 'common core' is taken to consist of a 'bare theory'—in itself understood to be a triple $\langle \mathcal{S}, \mathcal{Q}, \mathcal{D} \rangle$ of (respectively) states, quantities, and dynamics, isomorphic representations of which are contained in the structure of each of the dual theories under consideration (along with, potentially, further theory-specific structure). For my purposes, it suffices to know that the common mathematical core of two dual theories *can* be constructed in a well-defined manner.

[15]Via 'naïve interpretation', in the sense of footnote 8.

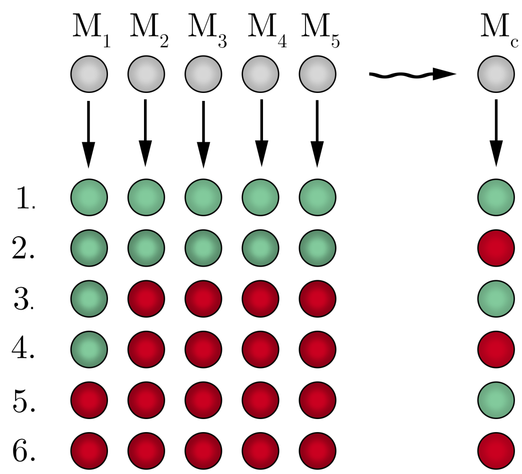

Figure 5.1: Five dual solutions $\mathcal{M}_1, \ldots, \mathcal{M}_5$, and their common mathematical core, $\mathcal{M}_c$. 'Naïvely interpreted', such solutions may be understood as representing certain worlds (to which they are 'isomorphic'—cf. footnote 8). If one such world is green, this indicates that this world is regarded as being a legitimate candidate for being the actual world. If the world is red, on the other hand, this indicates that this world is not regarded as being a legitimate candidate for being the actual world. Options (1)-(6) categorise different verdicts on which of these worlds are regarded as being candidates for being the actual world in this sense.

be regarded as being a legitimate candidate for being the the actual world, I colour it green; otherwise, I colour it red.

Clearly, a range of interpretative options are possible. If each of the worlds associated with $\mathcal{M}_1, \ldots, \mathcal{M}_5$ are regarded as being legitimate candidates for being the actual world (cases (1) and (2)), then we have a case of underdetermination: for given a set of empirical data compatible with one of these solutions, we do not know which of these five worlds is the actual world. Within this situation, two further options are possible: either the world represented by $\mathcal{M}_c$ is also a legitimate candidate for being the actual world (case (1)), or it is not (case (2)). I discuss the underdetermination approach in §5.5.

Suppose instead that just one of the original dual solutions is regarded as being a legitimate candidate for being the actual world (cases (3) and (4)); this is the discriminatory approach.[16] Within this scenario, two sub-scenarios again arise: either W_c is regarded as also being a legitimate candidate for being the actual world (case (3)), or it is not (case (4)). In §5.6, I reappraise the discriminatory approach.[17]

Finally, suppose that *none* of the worlds represented by $\mathcal{M}_1, \ldots, \mathcal{M}_5$ are regarded as being legitimate candidates for being the actual world (cases (5) and (6))—this is the nihilist gambit. Again, within such a scenario, either W_c may be regarded as itself being a legitimate candidate for being the actual world (case (5)), or it may not (case (6)). I discuss nihilism in §5.7.

[16]Clearly, the decision to regard $\mathcal{M}_1$ as representing this world in figure 5.1 is made without loss of generality.

[17]Let me demonstrate that (3) and (4) are distinct options by way of an analogy with a well-known case in the literature on symmetry transformations. Consider models of Newtonian gravitation theory, set in Newtonian spacetime (cf. [64, ch. 2]); let $\mathcal{M}_1, \ldots, \mathcal{M}_5$ correspond to solutions of this theory which differ only with regard to the absolute velocity of (the centre of mass of) the entire material content of the universe; let $\mathcal{M}_1$, in particular, be the model which states that the absolute velocity of (the centre of mass of) the entire material content of the universe is zero; and let $\mathcal{M}_c$ be the model of Newtonian mechanics set in Galilean spacetime corresponding to $\mathcal{M}_1, \ldots, \mathcal{M}_5$, in which the notion of absolute velocity has been excised. Note that $\mathcal{M}_1$ and $\mathcal{M}_c$ make different ontological claims about the world: the former states that the absolute velocity of the material content of the universe is zero; the latter states that this is undefined. (Cf. [210, pp. 45-46].)

With this in mind, consider (3) and (4), beginning with the latter. According to this position, only $\mathcal{M}_1$ of $\mathcal{M}_1, \ldots, \mathcal{M}_5$ is a legitimate candidate for representing the actual world. Why would one think this? One reason would be on the grounds of the principle of sufficient reason (PSR)—i.e. (roughly speaking) Leibniz's principle that God must have a sufficient reason to realise one of a class of symmetry-related models of a given theory (more on Leibniz's principles below). Maudlin articulates the point clearly, in the context of the same example:

> But on reflection, the PSR argument cannot get off the ground in the case of absolute velocity. Suppose that God wishes to create the material world in a heretofore empty absolute space. God could give the material world, as a whole, any absolute velocity in that space without affecting the relative positions and motions of bodies. Among all of these possible velocities, one stands out as special: absolute rest. For if God should give the material world any nonzero velocity, He would have to choose a *direction* in absolute space for that velocity to point. [137, p. 48]

If one accepts such reasoning, one may embrace option (4) in my taxonomy of options. In addition, however, there still remains the question whether $\mathcal{M}_c$—which, recall, represents (naïvely interpreted) a *distinct* world to $\mathcal{M}_1$, in which there is no meaningful notion of absolute velocity—is a legitimate candidate for representing the actual world. If one *does* think this, one will be pushed to (3)—one will think that either the universe is at rest, or that the actual world does not contain facts about absolute velocity at all. If one does *not* think that $\mathcal{M}_c$ is a legitimate candidate for representing the actual world (perhaps if one is strongly wedded to a notion of absolute velocity in order to maintain the coherence of one's metaphysics), one will remain with option (4).

Setting things up in the above manner is significantly more nuanced than extant taxonomies of interpretative options. For example, this framework illustrates that the discriminatory approach—represented in cases (3) and (4)—is not to be regarded as disjoint from the common core approach—represented by cases (1), (3), and (5).

After discussing options (1)-(6) in §§5.5-5.7, I consider in §5.8 whether there exist any *alternatives* to the common core approach. I identify two: (i) a tactic—common in the physics literature—of attempting to embed the two dual theories into some 'deeper' theory (§5.8.1); and (ii) the 'pluralist' strategy indicated above, in which the structures of all dual solutions are regarded as being *jointly* instantiated (§5.8.2).

5.4 Antirealism and structuralism

Before I discuss the above-outlined realist strategies for the interpretation of dualities, it is worth taking some time to present and set aside certain positions in the philosophy of science, which bear upon the interpretation of dualities, but which I will not consider further in this chapter.

The first class of such positions consists in 'antirealist' views, which (in some sense) deny that questions of 'ontology' are meaningful, or at least worthwhile. Stronger versions of antirealism are to be found in logical positivism and logical empiricism. Loosely, advocates of these views prioritise the observable, and claim that 'metaphysical' questions regarding the (unobservable) 'ontology' of the world are, strictly speaking, nonsensical.[18] A weaker view is the constructive empiricism of van Fraassen [210], according to which we should be agnostic about claims read off from our theories of science regarding the unobservable, and have credence only in the statements

[18] For some contemporary discussion of the logical positivist movement, and its transmutation into logical empiricism, see e.g. [40, 82, 86].

read off from our theories regarding observable phenomena.[19]

Since both strands of antirealism deny (in some sense) that it is of value to be concerned with answers to questions of ontology, I set them aside in the remainder of this chapter—for it is precisely such questions which I investigate, and which are regarded (at least *prima facie*) as being legitimate and worthwhile in the context of investigations into the ontological commitments of theories related by dualities.

As a halfway house between realism and antirealism, one may embrace a 'structural realist' thesis. There are many brands of structural realism—see e.g. [80,114,116, 235]. One version of structuralism of interest is the following: (i) the solutions of our best theories of physics are constituted formally by certain mathematical objects; (ii) we should take the world represented by that solution to be 'isomorphic' to the solution itself;[20] (iii) such a world may or may not be amenable to description in terms of extant ontological categories—but regardless of whether that be so, we should *still* believe that the solution of the theory in question describes some possible world.[21]

This brand of structuralism is certainly an interesting thesis, which merits much investigation. However, I elide further such discussion in this chapter, for two reasons. First (and to repeat), I am interested in saying something, in light of our best theories of physics (and especially in light of dualities), about *what the world is really like*; this does not appear possible in general on the above structuralist view. Second, I incline towards a more cautious attitude, according to which we may *only* regard a

[19]Thus, there is a sense in which the antirealism of van Fraassen is 'epistemological', whereas the antirealism of the logical positivists and empiricists is 'metaphysical'.

[20]Cf. footnote 8. Note that there are some subtleties here. First, such a claim can only strictly hold true for fundamental theories, in which we believe that the totality of the world is modelled by that solution of the theory in question. Second, such a claim effaces important issues regarding gauge redundancies—i.e., aspects of the formalism of a given theory which we do not believe to have representational capacities.

[21]Potentially the actual world, if the empirical substructures of that solution match the empirical data from the actual world.

particular solution of a particular theory as representing a possible world (rather than being mere mathematics) when we have in hand a clear picture of *what that solution is supposed to represent*. In this latter regard, see Ch. 6 for detailed discussion.

5.5 Underdetermination

I return now to the taxonomy of realist options for the interpretation of dualities presented in §5.3; I begin with the underdetermination view. One might not regard the putative underdetermination presented by dualities as being problematic. Indeed, according to the underdetermination interpretation, solutions of dual theories are in fact all legitimate candidates for representing the actual world—and we cannot ascertain *which* of the worlds associated with those dual solutions (naïvely interpreted) is the actual world.

Should one rest satisfied with the underdetermination interpretation? Arguably *no*, for in cases of underdetermination, it is impossible to ascertain which of a class of empirically equivalent worlds—all empirically adequate to the actual world—is, in fact, the actual world. Thus, underdetermination gives rise to a sceptical challenge: absent a means of determining which of a class of worlds is the actual world, *we have at hand no determinate picture of what the world is really like*.

One might seek to overcome such underdetermination in the following way: identify the 'common mathematical core' of the duality-related solutions under consideration, and take the actual world to be represented by this common mathematical core.[22] This, in effect, takes us from scenario (2) of figure 5.1, to scenario (1). Note, however, that in itself such a move is *insufficient* to resolve the underdetermination

[22] Such a position is common in the literature on dualities—see e.g. [48, 51, 100, 133].

under consideration. Indeed, there is a sense in which, absent further philosophical details, such a move has made the situation *worse*: we have, in effect, identified a *further* world which is empirically adequate to the actual world.

If such a 'common core approach' is to constitute a viable route to resolving the underdetermination, then it must be augmented by some *philosophical* reasoning, such that the original dual solutions do not in fact constitute legitimate descriptions of the actual world. One option here would be to appeal to putative super-empirical virtues of just one of the solutions under consideration (typically, though not necessarily, the common core solution $\mathcal{M}_c$)—e.g., simplicity, explanatory power, etc.—in order to break the underdetermination. However, there remains a gap to be bridged between such putative virtues, and truth-conduciveness (cf. [210, §4.1]). Only if such a bridge is constructed will appeal to such virtues move us from scenario (1) to scenario (3)/(5) of figure 5.1. I return to this issue of breaking the underdetermination in §5.6, on the discriminatory approach.

In this connection, one might think that another philosophical principle delivering scenario (3)/(5) from scenario (1), and thereby breaking the underdetermination, is Occam's razor—i.e., the principle that, all else being equal, otiose structure should not be introduced into one's ontology. Note, though, that as with the virtues discussed above, such a principle is first and foremost a *practical* principle, reminding us that it is (in general) preferable to work with (solutions of) more parsimonious physical theories.[23] Again, however, such a principle in itself does nothing to rule out as candidates for representing the actual world the other dual solutions under

[23]In fact, there is a sense in which this is not true in the case of dualities. For example, electromagnetism formulated in terms of a vector potential A^a is 'theoretically equivalent' (for my purposes: dual) to electromagnetism formulated in terms of the Faraday tensor F_{ab} (cf. [223,225,229]). Though the former theory has more structure (i.e., degrees of freedom) than the latter, and so is in this sense 'less parsimonious', there nevertheless exist many practical virtues of using the former over the latter—e.g. its amenability to variational principles; locality principles; etc. (cf. [56, 150]). Of course, strictly speaking this is compatible with Occam's razor, for in the foregoing language, all else is *not* equal.

consideration.

What is needed is a more robust metaphysical principle to exclude all but one of the dual solutions under consideration as being candidates for representing the actual world. One obvious option here is Leibniz's 'principle of the identity of indiscernibles' (PII), which (applied to worlds) can be understood (for our purposes) to state that there can be no 'distinctions without a difference'—that is, no empirically equivalent but physically distinct worlds, which vary with respect to unobservable structure.[24] Embracing such a metaphysical thesis seems to deliver us from scenario (1) to scenario (5)—since only W_c does not possess the variant undetectable structure under consideration, and is thereby compatible with this principle. Of course, however, the question naturally arises at this juncture: *why* should we commit ourselves to such a metaphysical principle?

5.6 Discrimination

In this section, I examine the discriminatory approach, according to which only one of the dual solutions under consideration is a legitimate candidate for representing the actual world. This approach is represented in scenarios (3) and (4) of figure 5.1. Claims appearing to be consonant with this approach are found in particular in the literature on the AdS/CFT correspondence. As Oriti points out, many string theorists speak as if the four-dimensional spacetime of the boundary theory is real, with the bulk spacetime appearing only as an auxiliary construction [158]. For example, Horowitz and Polchinski note that the AdS/CFT correspondence is a little different from other dualities in that the conformal field theory side is exactly understood,

[24] The PII was famously presented by Leibniz in the *Leibniz-Clarke Correspondence* [3]. For contemporary discussion of Leibniz's principles, and in particular the PII, see [188,191].

whereas the string theory side is only approximately understood. Building on this, they write,

> In the AdS/CFT case, the situation may not be so symmetric, in that for now the gauge side has an exact description and the string/gravity side only an approximate one: we might take the point of view that strings and spacetime are "emergent" and that the ultimate precise description of the theory will be in variables closer to the CFT form. [96, p. 230]

Here, Horowitz and Polchinski appear to claim that since one of a pair of dual theories (here, the conformal field theory) is better understood, we should privilege the ontological claims associated with solutions of that theory over those of its dual (here, the bulk string theory). Faced with such a passage, the question arises naturally: why should the *epistemological* fact about what human beings happen to currently know about two dual theories relative to one another warrant the *metaphysical* conclusion that the theory about which we currently know more must give the correct description of the world? Such worries have been expressed by Teh [204, §4], and most explicitly by Dieks, van Dongen, and De Haro [49, 59]. What is needed is some argument to the effect that one description of the world has metaphysical priority over its dual; such authors, however, treat this with suspicion—for example, Teh writes, "We have no good reason to think of the gravitational side of the duality as metaphysically emergent from the gauge theory side, or vice versa" [204, p. 310].

Is this response to Horowitz and Polchinski reasonable? In fact, there is perhaps room to defend the discriminatory apporach to dualities such as the AdS/CFT correspondence in the face of such criticism. Here is an alternative way to read Horowitz and Polchinski: it is not that we simply better *understand* one of the two dual theories (namely, the CFT); rather, it is that we only *have available* the full mathematical structure of the CFT, whereas the other dual theory—the AdS string theory—is only *partly constructed* (our understanding of the AdS side of the AdS/CFT duality is inherently perturbative—cf. [4]). The question here is not (*pace* the argument above) one of our only investing with ontological import those physical theories which we best understand, but rather one of our only investing with ontological import those physical theories which *which we actually have to hand*. Absent a completion of the AdS string theory, why take seriously its ontological claims (naïvely interpreted)—or even think that such a coherent completion and subsequent interpretation is to be had?[25]

Though it seems to me that such a defence of the discriminatory approach is reasonable in context of dualities such as the AdS/CFT correspondence, in which the full structure of only one of the duality-related theories is available, I consider now other ways in which the metaphysical primacy of one of the dual theories might be established. Such principles might be e.g. the super-empirical, or metaphysical, principles introduced in §5.5. However, as we have already seen, such super-empirical principles do not go far enough, for they do not *preclude* certain worlds from being legitimate candidates for being the actual world; moreover, while metaphysical principles such as the PII arguably do not face such difficulties, we appear to lack independent reason to embrace such principles.

It is worth dwelling a little longer upon how a defence of the discriminatory approach based upon the PII might proceed. Consider the situation in which two theories $\mathcal{T}_1$ and $\mathcal{T}_2$ are dual, with each solution of $\mathcal{T}_1$ corresponding to a class of solutions of $\mathcal{T}_2$.[26] If one also embraces the PII, then one should not regard the elements of this class of solutions of $\mathcal{T}_2$ as representing distinct worlds. But which (unique) world *should* one take these solutions to represent? One natural answer is that one should take

[25] There is some parallel between the caution advanced in the above paragraph, and that presented in the context of dualities in Ch. 6.

[26] Cf. footnote 3.

this world to be that represented by the unique solution of $\mathcal{T}_1$ (naïvely interpreted) to which this class of solutions of $\mathcal{T}_2$ corresponds.[27]

Note that here the PII is being put to a *different* use to that in §5.5. While in that case, the principle was used to establish that *neither* dual solution represents the actual world (rather, only their mathematical common core does so), here the principle is used to establish, in light of 'gauge redundancy' in one of the two dual theories under consideration, that the ontological claims of the *other* dual theory are to be preferred. These are, then, inter- versus intra-theoretic applications of the PII; in my view, both are in principle legitimate.

5.7 Nihilism

If all dual theories are to be considered on a par *vis-à-vis* the legitimacy of their ontological claims, then one must advocate either interpretations (1) and (2) in figure 5.1 (*viz.*, underdetermination interpretations), or interpretations (5) and (6). It is these latter two approaches which I now discuss; I call these 'nihilist' strategies. In particular, I focus on (6), for we have seen above circumstances in which one may be led to interpretation (5)—e.g., through embracing both the common core interpretation and the PII, the latter in order to exclude the legitimacy of the original dual solutions *qua* descriptions of the actual world.

As I see it, one might endorse nihilism for two reasons. The first applies in cases in which the dual theories under consideration are not expected to be final theories. In this case, one might claim that solutions of non-final theories (naïvely interpreted) simply cannot be understood to represent *any* possible world. There is something

[27] This, indeed, is exactly the standard answer given in the case of e.g. the theoretical equivalence of Newtonian gravitation theory set in Galilean spacetime, and Newton-Cartan theory—cf. [82,128,223].

to be said for this view, for consider e.g. solutions of Newtonian mechanics—an uncontroversial case of a non-final theory. It is well-known that the stability of matter cannot be accounted for in this theory; for this, one must proceed to some quantum-mechanical successor. But in that case, how *could* solutions of this non-final theory constitute legitimate candidates for representing the actual world? Thus, in this regard, one might be a nihilist about the ontological claims of any non-final dual theory, while still remaining a scientific realist—for in this case, one might still maintain that solutions of a *final* theory *could* constitute legitimate candidates for representing the actual world.

The second sense in which one might be a nihilist is the following. Suppose that, for antecedent reasons, one is unsympathetic to a particular research programme, e.g. string theory. In that case, one might reject (for said to-be-articulated antecedent reasons) solutions of all e.g. dual string theories as being legitimate candidates to describe the actual world, while remaining a realist, for one might think that solutions of *other* theories (e.g. loop quantum gravity, or extensions thereof) may legitimately describe the actual world.

In brief, then: nihilism is compatible with scientific realism just so long as there is *some* solution of *some* theory which one takes to constitute a legitimate description of the actual world. Perhaps such a solution is provided by looking to the mathematical common core of the dual solutions under consideration—in which case, one finds oneself in situation (5) of figure 5.1. Otherwise, one finds oneself in situation (6) of figure 5.1. But what solutions *could* describe the world, if not the common core, and not the original dual solutions under consideration? It is to this question that I now turn.

5.8 Alternatives to the common core

So far, I have considered underdetermination, discriminatory and nihilist approaches to the interpretation of dualities—these correspond respectively to options (1) and (2), (3) and (4), and (5) and (6) of figure 5.1. In this section, I consider whether there exist any alternatives to the 'common core approach' (options (1), (3), and (5) in figure 5.1). In my view, there exist (at least) two such options: (i) rather than construct a theory which represents the 'common core' of the original dual theories, construct a new theory in which each of the dual theories are *embedded* (§5.8.1); (ii) argue that each dual theory describes, correctly but partially, the same *one* world (§5.8.2).

5.8.1 Overarching theories

The common core approach purports to identify a possible world 'isomorphic' to the mathematical structure common to the dual solutions under consideration. Though this approach is popular in the philosophy of physics literature (see e.g. [48,100,131,133,184]), it is not the only live interpretative option purporting to break the underdetermination. Indeed, a distinct position—widely embraced in the physics community in the context of string-theoretic dualities—is to embed the spaces of solutions of the two dual theories under consideration into that of some deeper, 'overarching' theory.

What is meant by an 'overarching' theory? The answer to this question is best given by way of example. It is sometimes claimed that the five superstring theories are certain 'limits' of some deeper, 'M-theoretic' structure, in exactly the sense that their solutions spaces can be embedded into that of this deeper theory—cf. figure 5.2. Such an 'M-theory' is conceptually *distinct* from the common core of the dual string theories under consideration.

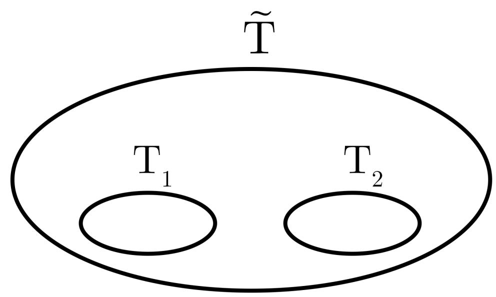

Figure 5.2: The spaces of two dual theories $\mathcal{T}_1$ and $\mathcal{T}_2$, embedded into the solution space of an 'overarching' theory, $\tilde{\mathcal{T}}$.

A parallel example of this manoeuvre can be drawn from the history of physics: consider the relation between Heisenberg matrix mechanics and the (putatively) empirically equivalent Schrödinger wave mechanics, in the 1920s and 30s.[28] Ultimately, the unification of these two theories did not proceed by finding their mathematical common core; rather, the two theories were embedded into a *deeper* theory—what we now refer to as orthodox quantum mechanics (see [146, §5]). This latter theory has a *richer* space of solutions than that of the two original theories—again, the structure is as per figure 5.2.

What to make of this approach with regard to the problem of underdetermination arising in the context of dualities? Merely embedding the spaces of solutions of the two dual theories into that of some 'deeper' theory does *not* in itself resolve the underdetermination—for again, given the empirical evidence compatible with one dual solution, it is not clear whether one should embrace the ontological claims of that solution, or of its dual, or of the overarching theory.

In fact, the situation here is even more subtle, and merits more detailed consideration; two scenarios are possible. First, if this 'embedding' of (the solution spaces of) two dual theories $\mathcal{T}_1$ and $\mathcal{T}_2$ into that of some overarching theory $\tilde{\mathcal{T}}$ is such that

[28] For a detailed study of the relations between these two early approaches to quantum mechanics, see [145, 146].

solutions of $\mathcal{T}_1$ and $\mathcal{T}_2$ *just are* solutions of $\tilde{\mathcal{T}}$, then it is clear that, for those solutions, $\tilde{\mathcal{T}}$ (naïvely interpreted) does not offer a distinct ontological picture over and above that of $\mathcal{T}_1$ and $\mathcal{T}_2$. In that case, we cannot interpret the ontology of these solutions in terms of the ontology of $\tilde{\mathcal{T}}$; that is, the introduction of $\tilde{\mathcal{T}}$ does nothing to resolve the problem of underdetermination which arises in the context of string-theoretic dualities. In such a case, one seeking this end would need recourse to (for example) one of the other interpretative options elaborated in this chapter.[29]

Second, suppose that (somehow, in some sense not fully captured by e.g. figure 5.2) solutions of $\tilde{\mathcal{T}}$ (naïvely interpreted) *are* taken to offer a distinct ontological picture from solutions of $\mathcal{T}_1$ and $\mathcal{T}_2$. In that case, one *may* say that the problem of underdetermination is resolved by interpreting solutions of $\mathcal{T}_1$ and $\mathcal{T}_2$ in terms of the ontology associated to the solutions of $\tilde{\mathcal{T}}$—however, as before, the problem is only truly resolved if some further, metaphysical principle is embraced—to the effect that worlds represented by solutions of $\mathcal{T}_1$ and $\mathcal{T}_2$ are *not* legitimate candidates for being the actual world, whereas those represented by solutions of $\tilde{\mathcal{T}}$ *are* legitimate such candidates. In this case, however, it is unclear that e.g. the PII could deliver this verdict—for it need not be the case that the solutions of $\tilde{\mathcal{T}}$ have *less* structure than those of $\mathcal{T}_1$ and $\mathcal{T}_2$ (as was the case in the common core approach)—but this is a crucial assumption of PII-style arguments, which are used to conclude that the ontology of the 'new' theory under consideration is to be preferred. Perhaps some *explanatory* thesis may be advocated here instead: 'since the overarching theory can account for more physical scenarios in its enriched space of solutions, that approach is to be preferred.' However, again, it is not clear whether touting of such super-empirical virtues truly resolves the *metaphysical* problem of underdetermination. Moreover, it is not clear that such explanatory reasoning is sound, for recall that in the context of classical gravity, the

[29] In this sense, the overarching theory approach can be combined with the common core approach. For further detailed discussion of the differences between these two approaches, and the extent to which they can be combined, see [50].

fact that the solution space of general relativity is enriched over that of the alternative programme of 'shape dynamics' is sometimes taken to constitute an advantage of the latter theory over the former.[30]

5.8.2 Pluralism

The second alternative to the common core approach is 'pluralism'.[31] On this approach, we consider the (distinct) structures of each of the dual solutions under consideration as describing *co-instantiated* structures in the actual world. That is, on this view, dual solutions may be taken to represent parts of *one* world. Each dual describes a seemingly different reality, but each of the structures under consideration represents a numerically distinct part of one world. Thus, according to pluralism, the physical world is in a certain sense fragmented.[32]

The pluralist strategy must pass a number of hurdles, if it is to be regarded as being successful. First, if pluralism is to resolve the putative underdetermination arising in cases of dualities, then, as before, some principled argument according to which the original dual solutions are *not* legitimate candidates for representing the actual world must be issued. Note, though, that this is not a problem *particular* to pluralism.

As I see it, there exist two central (and serious) issues for pluralism: an overdetermination problem, and an ontological problem. On the former, since each of the

[30]The claim here is that, since shape dynamics has a restricted space of solutions as compared with general relativity, it is 'more predictive'. See e.g. [9,138] for further discussion.

[31]Again: the idea for this position is due to Baptiste Le Bihan.

[32]Pluralism is incompatible with my definition of the common core approach, since according to the latter we should read the ontology in (roughly speaking) the intersection of the mathematical structures of the two duals under consideration, rather than the union of those structures as on the pluralist approach. Nevertheless, a 'fragmented' ontology could, in principle, be consistent with the common core approach.

individual dual solutions has certain empirical substructures, taken to correspond to the body of empirical data in the actual world, it seems that the pluralist strategy has, in a certain sense, swapped a problem of underdetermination for a problem of *over*determination, for now all of the dual structures may be taken to account for that body of empirical data. On the latter, the ontological problem is to understand what such a 'fragmented' world might look like, and whether the notion is consistent.

Let us focus upon the overdetermination problem. There exist two central options available by way of response to this issue: either the pluralist may claim (a) that only one of the co-instantiated dual structures accounts for the observed empirical data in the actual world, or she may claim (b) that that all such structures account *non-redundantly* for that data. In my view, there exist legitimate concerns regarding both proposals; let us discuss them in turn.

On the former view—(a)—the observed empirical data in the actual world may be accounted for by appeal to just *one* of the dual structures—though we do not know which one. Clearly, such an option is problematic, for it merely pushes the putative underdetermination arising in the case of dualities from the question of which of a number of worlds (i.e., those corresponding to the dual solutions, naïvely interpreted) could be the actual world, to the question of which of a number of distinct structures *within* a world could be that which accounts for the observed empirical data.

In order to maintain the latter view—(b)—the pluralist will argue that, in fact, all of the co-instantiated dual structures are *necessary* for accounting for the observed body of empirical data in the actual world.[33] It is not, however, clear that such a view is compelling, for it certainly appears that each of the dual structures could account for the observed empirical data *in and of themselves*. This point is best illustrated by

[33]One recent notion in metaphysics to which the pluralist might appeal here is that of 'multiple grounding'—cf. [44].

way of example.

Consider electromagnetism, formulated in terms of the vector potential A^a. It is well-known that the space of solutions of this theory partitions into equivalence classes of solutions, elements of each of which are related by a $\mathrm{U}(1)$ gauge transformation, and are regarded as being empirically equivalent (since each of these gauge-equivalent solutions gives rise—up to isomorphism—to the same Faraday tensor F_{ab}, which is taken to encode the observable data in the theory). In this case, we face an apparent problem of underdetermination analogous to that arising in the case of dualities,[34] for one might ask: 'which of the gauge-related solutions of electromagnetism instantiates the structure of the actual world?'

Of course, in this case it is standard to maintain that the 'true' ontology of solutions of the vector potential formulation of electromagnetism is that represented by the associated solution of the Faraday tensor formulation of electromagnetism.[35] Nevertheless, let us consider how the pluralist strategy would pan out in this case. On the analogue of the pluralist view here, *each* of the gauge-related solutions in a particular equivalence class of the theory, the empirical substructures of which correspond to the empirical data in the actual world, is instantiated in the actual world. But—and here is my response to this position—since each of these structures could *individually* give rise to that observed empirical data, it simply does not seem correct to state that all structures *together* must be co-instantiated in order to account for this data. To claim otherwise appears metaphysically otiose.[36] In my view, the central

[34]Indeed, one might treat the $\mathrm{U}(1)$ gauge symmetry of the vector potential formulation of electromagnetism as giving rise to a 'self-duality', in which case this example *just is* a case of underdetermination of the kind considered in this chapter.

[35]This is, in a sense, the analogue of the common core approach discussed above; for the (heterodox) analogue to the discrimination approach, see Maudlin's suggestion that there is "one true gauge" [136, p. 367].

[36]My mention of the co-instantiation of all gauge-related structures may remind the reader of Pitts' approach to gravitational energy in general relativity [160] (cf. Ch. 4). Exploring the overlap—if ultimately any—between this view and pluralism would constitute an interesting task for future pursuit.

challenge for the pluralist is to articulate a sense in which the above analogy does *not* hold, the co-instantiated structures *do* non-redundantly account for the observed empirical data, and therefore the problem of overdetermination is evaded.

Thus, pluralism faces a number of challenges, if it is ultimately to be regarded as being a compelling resolution to the problem of underdetermination in the case of dualities. Nevertheless, for the purpose of fully mapping the terrain on this topic, the view certainly deserves to be mentioned.

5.9 Conclusion

In this chapter, I have cut the issue of the interpretation of dualities along two distinct axes. First, which of the dual solutions under consideration should be taken to constitute a legitimate description of the actual world. The options here divide into three categories: underdetermination, discrimination, and nihilism. Second, candidate replacements for the ontology represented by the dual solutions under consideration, naïvely interpreted. Though the best-known approach in the philosophical literature in this regard is the common core approach, I have identified in this chapter two others: (a) appeal to an overarching theory; (b) what I have dubbed 'pluralism'. This latter position offers a novel avenue for the interpretation of dualities—albeit one that currently faces difficulties.

Chapter 6

Motivating Dualities

There exists a common view that for theories related by a 'duality', dual models typically may be taken *ab initio* to represent the same physical state of affairs, i.e. to correspond to the same possible world. In this chapter, I question this view, by drawing a parallel with the distinction between 'interpretational' and 'motivational' approaches to symmetries.

6.1 Introduction

The phenomenon of 'duality' is pervasive in theoretical physics—particularly string theory. As discussed in Ch. 5, I take in this thesis two physical theories to be dual when there exists an isomorphism between their spaces of dynamically possible models, such that models related by that isomorphism are empirically equivalent. According to a common view in the philosophical literature, duality-related models typically may be construed *ab initio* as representing the same physical state of affairs, i.e. as corresponding to the same possible world—in which case duality-related models are not only empirically equivalent, but also physically equivalent. Two motivations for this view are often advanced:

(1) This view of dualities aligns with a general conception of the philosophical import of *symmetry transformations*—namely, that models related by a symmetry transformation typically may be understood *ab initio* as being physically equivalent.

(2) This view of dualities accords with a perceived consensus within the contemporary theoretical physics community.

In this chapter, I question both (1) and (2). On (1), I deny that dual models may be regarded as physically equivalent absent a coherent explication of the common ontology underpinning this physical equivalence; and by the same token, I deny that symmetry-related models may be regarded as physically equivalent in the absence of such an explication. Thus, I argue for a reconstrual of the import of dualities and symmetries: dualities invariably at most *motivate one to seek* an understanding of how it is that dual models are to be regarded as physically equivalent; and by the same token, symmetries also invariably at most *motivate one to seek* an understanding how it is that symmetry-related models are to be regarded as physically equivalent. On (2), I cite a variety of evidence from the physics literature which calls into question whether this perception of such a consensus is correct.[1]

The format of this chapter is as follows. In §6.2, I recall again some of the central features of the semantic approach to scientific theories. In §6.3, I introduce Møller-Nielsen's distinction (drawn at [142, §2]) between the 'interpretational' and 'motivational' approaches to symmetries, using Newtonian gravitation theory as an illustrative example; I go on to defend the motivational approach.[2] In §6.4, I present no-

[1] In this regard, I follow the methodology of [20].

[2] In this regard, this chapter may be viewed as continuous with [142], offering further reasons to

tions of 'underdetermination' and 'theoretical equivalence' which will prove useful in my subsequent discussion of dualities in §6.5—in which I introduce a distinction between the interpretational and motivational approaches to *dualities*, and (again) defend the latter approach. Finally, in §6.6 I assess the extent to which the interpretational approach—a common view in the philosophical literature—is embraced in the theoretical physics community.

6.2 Models and Gauge

As discussed in previous chapters, on the semantic conception of scientific theories, a theory is associated with a class of models.[3] For a given theory $\mathcal{T}$, I take the most general class of associated models to be that of 'kinematically possible models' (KPMs) $\mathcal{K}$, which consists in tuples of specified geometrical objects. For example, the KPMs of general relativity (GR) are picked out by all triples of the form $\langle M, g_{ab}, \Phi \rangle$, where M is a four-dimensional differentiable manifold; g_{ab} is a Lorentzian metric field on M; and Φ is a placeholder for the matter fields of the theory.

Classically, a theory $\mathcal{T}$, with KPMs $\langle M, O_1, \dots O_n \rangle$ (where the O_i are geometrical objects), comes with a set of dynamical equations for the O_i. The KPMs of $\mathcal{T}$ in which the O_i obey those dynamical equations form a subset $\mathcal{D} \subset \mathcal{K}$, the 'dynamically possible models' (DPMs) of $\mathcal{T}$. For example, in the case of GR, only those triples $\langle M, g_{ab}, \Phi \rangle$

endorse the motivational approach, as well as providing an extended application of the interpretation/motivation distinction to the case of dualities.

[3]One should distinguish the claim that a given theory has an associated class of models from the (more controversial) claim that a theory should be *identified* with such a class of models. In this thesis, I embrace the former, but remain agnostic on the latter.

the geometrical objects of which satisfy the Einstein field equations

$$G_{ab} = 8\pi T_{ab} \tag{6.2.1}$$

—the dynamical equations of the theory, which relate g_{ab} to the stress-energy tensor T_{ab} of the Φ—in *addition* to the dynamical equations of the Φ, are DPMs. Quantum mechanically, the story changes: some of the O_i in the KPMs of $\mathcal{T}$ are understood to be *operator-valued*; DPMs are picked out as those KPMs the geometrical objects of which satisfy certain correlation functions (for relevant aspects of the structure of quantum field theory, see e.g. [61,159,197]; for further philosophical details regarding the above approach, see [175, ch. 5]).

Models of a theory $\mathcal{T}$ are interpreted as representing possible worlds. Sometimes, however, we may wish to interpret two or more distinct models as representing the *same* world. In that case, the space of KPMs $\mathcal{K}$ of $\mathcal{T}$ is partitioned into classes of 'gauge-equivalent' models—which are interpreted as representing the same world—and the multiplicity of models representing the same world is an example of a 'gauge redundancy'.[4] In the case in which the interpretation of $\mathcal{T}$ leads to gauge redundancy, one may construct a reduced space of models $\tilde{\mathcal{K}}$, in which gauge-related models are mathematically identified.[5] This in turn induces a reduced space of DPMs, $\tilde{\mathcal{D}} \subset \tilde{\mathcal{K}}$.[6]

[4]It should be stressed that the term 'gauge redundancy' is deployed in this thesis in a broader sense than that typically found in the physics literature, where the term is often reserved for certain 'internal' symmetries associated with Yang-Mills type theories. For philosophical discussion, see e.g. [92, 216, 225].

[5]For a concise expression of these points in the language of category theory, see [223, 225, 229].

[6]One assumes that two models cannot be gauge-equivalent if they satisfy different dynamics. While one might worry that this understanding of gauge redundancies effaces the possibility that two models with different dynamics may correspond to the same possible world (and thereby pose problems for the interpretation of *dualities*—see §6.4 below), this is not correct, for nothing in the above precludes the possibility that there exist other relations which may allow for *inter*-theoretic model identification.

6.3 Interpretation and Motivation

6.3.1 Two Approaches to Symmetries

The above is purely formal; there remains an outstanding question concerning *when* two models of $\mathcal{T}$ should be interpreted as representing the same possible world. One popular line (found—although not necessarily endorsed—in e.g. [8, 18, 38, 47, 54, 60, 89, 104, 154, 190, 232]) is the following: two models of $\mathcal{T}$ typically may be regarded *ab initio* as representing the same possible world when they are related by a *symmetry transformation*—even absent a coherent explication of their shared ontology.[7,8]

According to this 'interpretational' approach—first articulated by Møller-Nielsen in [142, §2]—in the presence of symmetry-related models, we are (a) typically warranted in interpreting those models as representing the same possible world—even absent a coherent explication of their common ontology; then may (but are not required to) go on to (b) identify such models, to construct a reduced space of KPMs $\tilde{\mathcal{K}}$; and finally (c) seek to explicate the ontology of the models of $\tilde{\mathcal{K}}$. This is in contrast with the 'motivational' approach [142, §2], according to which the existence of symmetry-related models first (a) motivates us to provide an explication of the shared ontology of these models; but only once such an explication is forthcoming should we (b) interpret those models as representing the same possible world; and (potentially) (c) identify those models to construct a reduced space of KPMs, $\tilde{\mathcal{K}}$.[9]

[7]Clearly, such a claim has substance only once an appropriate definition of a 'symmetry transformation' is provided; this matter is addressed in detail below.

[8]What is meant by such an explication will be made explicit over the following subsections. This explication must cohere both internally, and with the structure of the models under consideration (it is, therefore, insufficient to simply assert that the two models under consideration be interpreted as corresponding to some *arbitrary* possible world).

[9]Here, I say *potentially*, for there does not necessarily exist any pressure to construct such a $\tilde{\mathcal{K}}$. To illustrate, consider the case of models related by a hole diffeomorphism in GR: even one who interprets such models as corresponding to the same possible world is not obliged to construct such a reduced theory. Cf. §6.3.2.6 below.

Why 'typically', in the above presentation of the interpretational view? A supporter of this view *may* impose certain further criteria for when symmetry-related models are to be regarded as physically equivalent—and so need not always *actually* interpret such models as being physically equivalent. For example, even for the interpretationalist it is plausible that not all symmetry-related models should be interpreted as corresponding to the same possible world, for consider e.g. the case of *Galileo's ship*, in which only a subsystem in a model of Newtonian mechanics is boosted—in this case, we have two symmetry-related models, which nevertheless clearly do *not* correspond to the same possible world.[10] Inserting this 'typically' clause does not obscure the interpretational view, however—for the salient point is the following: on the interpretational approach, the decision to interpret symmetry-related models as being physically equivalent *need not wait upon an explication of their shared ontology*.[11]

Clearly, if the above 'interpretational' claim, and its 'motivational' alternative, are to have substance, an appropriate definition of a symmetry transformation must be provided. Suppose first that one defines such a transformation to be one upon the O_i in the KPMs of any given theory $\mathcal{T}$, such that DPMs of $\mathcal{T}$ are always taken to DPMs. Such a definition clearly will not do, for, as Belot points out, it is much too broad: [19, p. 6]

> Ordinarily, symmetries of theories are hard to come by. But some remarkable theories have atypically large symmetry groups. The definition above effaces this sort of distinction between theories. For if we allow arbitrary permutations of the solutions of a theory to count as symmetries, then the

[10] For some discussion of such issues, see [48, 89, 176]; in this chapter (modulo some brief considerations in §6.5.3), I set these complications aside by considering only symmetry transformations which act 'globally' upon the O_i of KPMs of $\mathcal{T}$, rather than upon proper subsystems in those models.

[11] My thanks to Neil Dewar for helpful discussion on this point.

size of a theory's group of symmetries depends only on the size of its space of solutions.

Given this, a more nuanced definition of a symmetry transformation is required; following e.g. [38, 45, 104], in this chapter I take this to be one of *empirical equivalence*. Accordingly, I define a symmetry transformation as follows: a symmetry of a theory $\mathcal{T}$ is any automorphism of the space of DPMs of $\mathcal{T}$, *such that models related by that transformation are empirically equivalent*.[12] By 'empirical equivalence', I mean in turn that all the structures in the models under consideration corresponding to 'physically observable data' are identical between those models—that is, that the 'empirical substructures' of these models in the sense of van Fraassen [210, p. 64] coincide.[13]

It is important to be clear that I am not *endorsing* the above epistemic definition of a symmetry transformation (or the parallel epistemic definition of a duality presented in §6.4.1). Rather, I am merely taking it as given in this chapter that symmetry-related models are empirically equivalent, while bracketing questions such as (i) whether that criterion should constitute part of the 'correct' definition of a symmetry transformation; and (ii) whether it is universally true that symmetry transformations relate (all and) only empirically equivalent models. It is, however, worth noting that for models to even potentially be physically equivalent, they must at the very least be empirically equivalent. Thus, even if one rejects the above definition of 'symmetry', one should recognise that the 'symmetries' *relevant* to my discussion here will satisfy the condition of being empirically equivalent. For a further critical discussion of these and related issues, see [143].

[12] This definition of a symmetry transformation has the merit of being broadly analogous with my construal of dualities, presented in Ch. 5 and §6.4.1.

[13] Each of [38, 45, 104] offer more nuanced ways of cashing out the 'empirical equivalence' criterion in the above definition of a symmetry transformation—for example, Dasgupta appeals both to a notion of 'how things look' [45, §6.3], and to Quinean 'observation sentences' [45, §6.3] (for details of such observation sentences, see [173, 174]).

On the first definition of a symmetry transformation above—*viz.*, that considered and dismissed by Belot—not only would it be incorrect to interpret *ab initio* all symmetry-related models as being physically equivalent (for then all models of the theory in question would be afforded the same interpretation), but, moreover, one would clearly *not even be motivated* to find an interpretation according to which such models are physically equivalent. Thus, the motivational approach is incompatible with such a definition of a symmetry transformation. *Prima facie*, neither of these points holds for the revised definition of a symmetry transformation, featuring the additional criterion of empirical equivalence. The reason for this is that such a definition is more restrictive—so it *might* be the case that one can argue that all symmetry-related models may be regarded *ab initio* as being physically equivalent, in line with the interpretational approach; moreover, one *is* apparently motivated to find a coherent interpretation according to which such models are physically equivalent, essentially on the grounds of Occam's razor: since any structure leading to such models being interpreted as physically distinct would not be part of the empirical substructures of those models (which are identical), that structure is variant yet undetectable—so we have good *prima facie* grounds for seeking to excise it.

6.3.2 Newtonian Gravitation Theory

In order to clarify and develop further the distinction between the interpretational and motivational approaches to symmetries, I consider in this section the case of Newtonian gravitation theory (NGT).[14] In §§6.3.2.1, 6.3.2.2 and 6.3.2.3, I introduce (respectively) the KPMs, fundamental interpretational postulates, and DPMs of NGT. In §6.3.2.4, I introduce three important classes of symmetries of NGT, before in §§6.3.2.5 and 6.3.2.6 discussing the interpretational and motivational approaches in the context

[14]For rigorous presentations of this theory, see e.g. [64, 81, 128, 170].

of these classes of symmetries.

Those readers uninterested in the technical details of NGT are advised to skip straight to §6.3.2.4. In my view, it is necessary to spell out the technical details of this theory, because I seek to provide a fully worked out example of what it *means* to fully explicate symmetry-related models' underlying ontology. More specifically, in my view the best way of conveying what a 'full explication' (or 'transparent understanding') of the reality underlying symmetry-related models amounts to is *by analogy*. Hence, the relevant technical details of this example should be spelled out in full, even though the basic ideas can plausibly be understood without them. (There is some overlap in the following with Ch. 3; however, the material presented here is at a higher level of detail, and used to make a distinct philosophical point.)

6.3.2.1 Kinematically Possible Models

In its field-theoretic formulation, KPMs of NGT (set in Newtonian spacetime—see [64, pp. 33ff.], and discussion below) are picked out by tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho \rangle$, where M is a four-dimensional differentiable manifold; t_{ab} is a temporal 'metric' field on M of signature $(1, 0, 0, 0)$; h^{ab} is a spatial 'metric' field on M of signature $(0, 1, 1, 1)$;[15] ∇_a is a derivative operator on M; σ^a is a vector field; and φ and ρ are scalar fields that represent the gravitational potential field and matter density, respectively. At the level of KPMs, the following four conditions hold:

[15]Strictly, neither t_{ab} nor h^{ab} is a metric field—see e.g. [128, p. 250]. Insofar as they are not metric fields, t_{ab} and h^{ab} are still tensor fields of rank $(0, 2)$ and $(2, 0)$, respectively.

$$h^{ab}t_{ab} = 0, \tag{6.3.1}$$

$$\nabla_a t_{bc} = 0, \tag{6.3.2}$$

$$\nabla_a h^{bc} = 0, \tag{6.3.3}$$

$$t_{ab}\sigma^b \neq 0. \tag{6.3.4}$$

I refer to (6.3.1) as an 'orthogonality' condition, and (6.3.2) and (6.3.3) as 'compatibility' conditions. (6.3.4) ensures that σ^a has a component in the temporal direction,[16] and so that the images of its integral curves may be used to represent the persisting points of absolute space.

6.3.2.2 Interpretative Principles

Following Malament [128, p. 252], I now introduce the following interpretive principles in NGT. Let I be an open interval in $\mathbb{R}$. Then, for all smooth curves $\gamma : I \to M$:

- γ is timelike[17] if its image $\gamma[I]$ could be the worldline of a point particle.
- γ can be reparameterised so as to be a timelike geodesic (with respect to ∇_a) iff $\gamma[I]$ could be the worldline of a free point particle.
- Clocks record the t_{ab}-length of their worldlines.

[16] I.e. is timelike, in the sense of footnote 17.

[17] Given any vector θ^a at a point $p \in M$, we can take its 'temporal length' to be $\left(t_{ab}\theta^a\theta^b\right)^{1/2}$. I further classify θ^a as either 'timelike' or 'spacelike', depending on whether its temporal length is positive or zero, respectively. I understand a smooth curve to be 'timelike' (respectively 'spacelike') if its tangent vectors are of this character at every point along the curve. Note that (6.3.4) ensures that σ^a is a timelike vector field.

If a particle has the image of a timelike curve as its worldline, then we call the tangent field ξ^a of that curve the 'four-velocity' field of the particle, and call $\xi^b \nabla_b \xi^a$ its 'four-acceleration' field (note that strictly this is a spacelike quantity, representing the instantaneous rate of change of the three-velocity of the body in question, as determined by an inertial observer). If the particle has a mass m, then its four-acceleration field satisfies

$$F^a = m\xi^b \nabla_b \xi^a, \tag{6.3.5}$$

where F^a is a spacelike vector field (on the image of its worldline) that represents the net force acting on the particle. This is the generalised form of Newton's second law for NGT.

6.3.2.3 Dynamically Possible Models

With these principles in mind, I am now in a position to make explicit the DPMs of NGT.[18] In NGT, one first imposes flatness of ∇_a via the field equation

$$R^a{}_{bcd} = 0. \tag{6.3.6}$$

A second field equation of NG is Poisson's equation,

$$h^{ab} \nabla_a \nabla_b \varphi = 4\pi\rho. \tag{6.3.7}$$

[18]One may question whether these laws faithfully represent Newton's thinking on these matters, since they make no reference to the persisting point of absolute space, as picked out by σ^a. For an arguably less anachronistic presentation of the laws of NGT set in Newtonian spacetime, see [170, §4.4]. The presentation of the dynamical laws of this subsection will suffice for the purposes of this chapter.

Finally, the gravitational force on a point particle of mass m is given by $-mh^{ab}\nabla_b\varphi$. It follows from (6.3.5) that if the particle is subject to no forces except gravity, and if it has four-velocity ξ^a, then it satisfies

$$-\nabla^a\varphi = \xi^b\nabla_b\xi^a. \tag{6.3.8}$$

6.3.2.4 Symmetries of Newtonian Gravitation Theory

The above presentation of NGT in hand, consider now the symmetries of this theory. The symmetry group of NGT includes three kinds of transformations that are worth singling out: (A) the 'static shift', which involves a time-independent translation of the total matter content of the original solution; (B) the 'kinematic shift', which involves a time-independent velocity 'boost' of the total matter content of the original solution; and (C) the 'dynamic shift', which involves a time-dependent translational acceleration of the total matter content of the original solution, plus an appropriate transformation of the gravitational potential field.[19]

It is possible—and useful—to characterise all of these symmetries model-theoretically. Taking our original model to be $\mathcal{M} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho\rangle$, a static-shifted model can be written $\mathcal{M}_{\text{stat}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi, d^*\rho\rangle$, where d is the appropriate diffeomorphism corresponding to a spatial translation. Straightforwardly—or 'literally'—understood, the world represented by $\mathcal{M}_{\text{stat}}$ differs from that represented by $\mathcal{M}$ with regard to which particular points of space are underlying various parts of the matter fields. For instance, if $\mathcal{M}$ represents the centre of mass of the universe[20] as being lo-

[19] The terms 'static shift' and 'kinematic shift' are relatively standard in the literature, and are originally due to Maudlin [135, §3]. The term 'dynamic shift' is slightly less standard, and is due to Huggett [97, §8.3].

[20] One worry regarding speaking of the 'centre of mass of the universe' is the following: this notion may only be well-defined under a certain restricted set of circumstances (for example, when the mass

cated *here*, then $\mathcal{M}_{\text{stat}}$ will represent the centre of mass of the universe as being located e.g. 3m *to the left of here*.

Consider now the kinematic shift. The generic model yielded by applying the kinematic shift to $\mathcal{M}$ can be written $\mathcal{M}_{\text{kin}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi, d^*\rho \rangle$, where d is now the appropriate diffeomorphism corresponding to a velocity boost. Straightforwardly understood, the world represented by $\mathcal{M}_{\text{kin}}$ differs from that represented by $\mathcal{M}$ with regard to the absolute velocity of the material universe. For instance, if $\mathcal{M}$ represents the centre of mass of the universe as being *absolutely at rest*, then $\mathcal{M}_{\text{kin}}$ will represent it as moving e.g. 3ms^{-1} *due North*.

Finally, the generic model yielded by applying the dynamic shift to $\mathcal{M}$ can be written $\mathcal{M}_{\text{dyn}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi', d^*\rho \rangle$, where d is a diffeomorphism corresponding to an element of the so-called 'Maxwell group' of transformations, and where the gravitational potential field is transformed by an appropriate 'internal' transformation.[21] Thus, straightforwardly understood, the world represented by $\mathcal{M}_{\text{dyn}}$ differs from that represented by $\mathcal{M}$ with regard to what the absolute translational acceleration of the material universe is alleged to be. For instance, if $\mathcal{M}$ represents the material universe as being *absolutely non-accelerating*, then $\mathcal{M}_{\text{dyn}}$ will represent its centre of mass as accelerating in a straight line under a gravitational force-field, at e.g. 3ms^{-2} *due North*.

In sum: the symmetries of NGT include transformations that map DPMs to other DPMs that *prima facie* represent physically distinct worlds. Nevertheless, no observer 'embedded' in any of these worlds can determine which world is hers: the worlds

density ρ is asymptotically zero at infinity). Given this, it may be preferable to resort to the following fix: use instead the centre of mass of some arbitrary body of matter. My thanks to Neil Dewar for raising this point.

[21] Following [64, §2.3], the Maxwell group of transformations is defined as $\vec{x} \to \vec{x}' = \mathbf{R}\vec{x} + \vec{a}(t)$; $t \to t' = t + d$. The 'internal' transformation on φ is defined as $\varphi \to \varphi' = \varphi - \vec{x} \cdot \ddot{\vec{a}} + f(t)$. For further details, see [112].

represented by these models are 'empirically indistinguishable'—so these symmetry-related models are indeed 'empirically equivalent', in line with the definition of a symmetry transformation presented in §6.3.1. This is because all *relative* distances and velocities between material systems are preserved among the worlds in question, and all an observer has empirical access to are (ratios of) such distances and velocities.[22] Thus, such an observer would not be able to determine whether she is stationary, moving uniformly, or accelerating relative to the persisting points of absolute space: all of these scenarios are underdetermined by the empirical phenomena.

6.3.2.5 Interpretation and Motivation in Newtonian Gravitation Theory

So much for the symmetries of NGT. How do the interpretational and motivational approaches discussed in §6.3.1 play out in this theory? In the present context, the distinction can be stated easily. Consider again the models $\mathcal{M}$, $\mathcal{M}_{\mathrm{stat}}$, $\mathcal{M}_{\mathrm{kin}}$, and $\mathcal{M}_{\mathrm{dyn}}$. According to the former view, it is legitimate to take *ab initio* all of these models—which *prima facie* represent distinct physical scenarios—to in fact represent the same state of affairs, i.e. the same possible world, even absent a coherent picture of their common ontology.

The motivational view, on the other hand, denies that it is permissible to so regard symmetry-related models as being physically equivalent. Rather, on this view, the symmetries of a theory invariably at most *motivate one to seek* a clear understanding of the common ontology underpinning such models' physical equivalence. That is, according to this view, models related by a symmetry transformation cannot be regarded as physically equivalent simpliciter. Instead, construing such models as physically equivalent is only justified once one has a clear understanding of the re-

[22] By 'ratios of' distances and velocities, I have in mind such notions relative to a pre-defined standard of measurement—e.g. the Parisian 'metre rod'.

ality allegedly underlying them: a clear understanding that we are, according to the motivational view, invariably motivated to seek by the symmetry in question. Thus, on the motivational view, absent a clear understanding of *how it could be* that $\mathcal{M}$, $\mathcal{M}_{\mathrm{stat}}$, $\mathcal{M}_{\mathrm{kin}}$, and $\mathcal{M}_{\mathrm{dyn}}$ are to be regarded as physically equivalent (in terms of a clear explication of their common ontology), we may not regard them as so being physically equivalent, i.e. as corresponding to the same possible world.

6.3.2.6 Mathematical Reformulation

Importantly, and as Møller-Nielsen stresses [142, §3], the motivational approach is *not* committed to the view that whenever one is presented with a theory $\mathcal{T}$ with a symmetry between mathematically distinct models, one is motivated to *mathematically reformulate* $\mathcal{T}$ so as to remove any such (alleged) representational redundancy (such that models of the reformulated theory are constructed by *quotienting* the space of models of the original theory by the action of the symmetry in question).[23] Rather—and this will become important in the discussion of dualities in §6.5—such a mathematical reformulation is motivated only when the models in question are *not* isomorphic, i.e. when (straightforwardly understood) they differ more than merely with regard to which objects play which qualitative roles.[24,25] As I will now discuss, this means that, according to the motivational view, in the case of NGT only the kinematic and dynamic shifts motivate us to mathematically reformulate the theory so as to remove any representational redundancy.

[23] One may here understand 'mathematical reformulation' to mean: an alteration of the space of models of the theory (whether KPMs or DPMs). This will become clear through the examples presented in this subsection.

[24] Here, 'object' refers to any substructure of the model in question—rather than (necessarily) to the geometric objects O_i introduced in the KPMs of a generic theory $\mathcal{T}$ in §6.2.

[25] Note that isomorphism of two spaces of models (e.g. $\tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{D}}_2$, associated respectively to two theories $\mathcal{T}_1$ and $\mathcal{T}_2$)—as introduced in §6.1, and discussed further in §6.4.1 below—should not be confused with isomorphism of a given pair of models themselves. It is the latter that is under consideration here.

The static shift in NGT is crucially distinct from the kinematic and dynamic shift—for the reason that the models in question in this case are isomorphic. And indeed, there exists a straightforward means of understanding such isomorphic models' physical equivalence, which necessitates no mathematical reformulation of the theory. This view goes by a variety of names in the literature: in spacetime contexts, it is most commonly referred to as 'sophisticated substantivalism'.[26] The sophisticated substantivalist denies that spacetime points possess primitive transworld identities; instead, they are 'contextually individuated' [115, §5]: they are not to be construed as being anything less, or more, than 'nodes' in the relational, geometrical structures in which they are embedded. This view is still a version of substantivalism, in the sense that it is committed to points of space being fundamental, basic elements of reality. Crucially, however, this view denies that there are any primitive, singular (haecceitistic) facts about spacetime points (e.g. *this particular* point of space is materially occupied) which would even allow for a physical distinction between statically shifted scenarios to be drawn.

Analogous considerations apply in the context of more modern physical theories; the diffeomorphism invariance of GR provides a case in point. Just as for the static shift in NGT, the existence of this symmetry is alleged to commit the substantivalist to a plurality of physically distinct, but nevertheless empirically indistinguishable, possibilities. Once again, we can phrase this in model-theoretic terms: taking a generic DPM of GR, $\mathcal{M} = \langle M, g_{ab}, \Phi \rangle$, we can apply an arbitrary diffeomorphism d to yield a new DPM, $\mathcal{M}' = \langle M, d^* g_{ab}, d^* \Phi \rangle$. A popular allegation—the canonical version of which can be found in [65, §4]—is that the spacetime substantivalist is committed to regarding the two worlds represented by these models as differing with regard to which particular points of the spacetime manifold are underlying various parts of the

[26]See, e.g., [169, p. 575]. Other names for this view include 'moderate structural realism' about spacetime [69, pp. 31-2] and 'non-reductive relationalism' [190, §5].

metric and matter fields.[27]

It should be clear that, if adopting sophisticated substantivalism constitutes a legitimate response to the alleged problem of NGT's static shift symmetry, it should count as an equally legitimate response to the alleged problem of GR's diffeomorphism symmetry. That is, adopting sophisticated substantivalism should be sufficient for one to be able to understand, in a perfectly transparent way, how it is that diffeormorphism-related models in GR are to be regarded as physically equivalent, without any mathematical reformulation of the theory being necessitated—just as in the case of the shift symmetry of NGT.[28]

Now return to the case of non-isomorphic symmetry-related models in NGT, namely $\mathcal{M}$, $\mathcal{M}_{\text{kin}}$, and $\mathcal{M}_{\text{dyn}}$. 'Literally understood', such models do not represent possible worlds which differ merely haecceitistically. Hence, adopting sophisticated substantivalism is by itself insufficient to be able to understand how such models are to be regarded as physically equivalent.[29] Thus, we are motivated to mathematically reformulate the theory so as to obtain a coherent understanding of the common ontology underpinning such models' physical equivalence.[30]

Such a mathematical reformulation of the theory is indeed possible. In fact, for the kinematic shift, it is trivial: one simply excises σ^a from KPMs of the theory—so that the question of two otherwise-identical models differing only in the absolute velocity

[27]Here I ignore the related (but distinct) 'indeterminism' objection to substantivalism in the context of GR raised at [65, §5]. The reasons for this are twofold. First, this objection is not directly related to the static shift argument in NGT. Second, sophisticated substantivalism also seems sufficient as a response (for more on this latter point, see [167, §4.1.4]).

[28]Cf. footnote 9. Of course, sophisticated substantivalism constitutes just one of many positions available in the vicinity of discussions of the hole argument. For a recent review of the literature, see [169, §7].

[29]For the parallel point in the case of dualities, see [176, §5.3]. Cf. §6.5.1.

[30]Note that, since the mathematically reformulated theory will have a different space of models to the original theory (cf. footnote 23), it may best be regarded as a *new* theory, distinct from the original (on the setup of §6.2).

of the centre of mass of the matter content they represent does not arise. In other words, one moves from Newtonian spacetime—where the persistence of points of space through time is assured (since, recall, the histories of these points are associated with the integral curves of the σ^a field), and where the associated notion of absolute velocity is physically meaningful—to Galilean spacetime, where the persistence of points of space through time and the associated notion of absolute velocity no longer make physical sense, but where the difference between straight (inertial) and curved (accelerating) trajectories through spacetime remains physically meaningful.[31]

In the case of the dynamic shift, reformulation is also possible—though somewhat less straightforward. Here—having already eliminated σ^a from the models of the theory—one replaces the flat derivative operator ∇_a of NGT with a (partly) dynamical derivative operator $\hat{\nabla}_a$,[32,33] for which the DPMs then require that the associated curvature tensor $\hat{R}^a{}_{bcd}$ satisfies

$$\hat{R}_{bc} = 4\pi\rho t_b t_c, \tag{6.3.9}$$

$$\hat{R}^a{}_b{}^c{}_d = \hat{R}^c{}_d{}^a{}_b, \tag{6.3.10}$$

$$\hat{R}^{ab}{}_{cd} = 0, \tag{6.3.11}$$

[31]For further discussion, see e.g. [64, §2.4] and [137, pp. 54-66]. Although such a reformulation of NGT may appear trivial from a modern four-dimensional, differentio-geometric perspective, it certainly would not have appeared so to Newton or his contemporaries. This appearance of triviality is arguably reinforced by the fact that, in setting up NGT, I have (following the canonical literature on this subject, in particular [81, pp. 71-94]) formulated the laws directly in terms of ∇_a, rather than σ^a. For more on this point, see [170, p. 134]; for a discussion of NGT which puts particular emphasis on the non-triviality of the move to Galilean spacetime, see [137, pp. 54-66].

[32]$\hat{\nabla}_a$ is related to ∇_a by $\hat{\nabla}_a = (\nabla_a, C^a{}_{bc})$, with $C^a{}_{bc} = -t_b t_c \nabla^a \varphi$; bracket notation for derivative operators means that $\hat{\nabla}_a$, ∇_a, and $C^a{}_{bc}$ are related by $\left(\hat{\nabla}_c - \nabla_c\right) \alpha^{a_1 \dots a_r}_{b_1 \dots b_s} = \alpha^{a_1 \dots a_r}_{d b_2 \dots b_s} C^d{}_{cb_1} + \dots + \alpha^{a_1 \dots a_r}_{b_1 \dots b_{s-1} d} C^d{}_{cb_s} - \alpha^{d a_2 \dots a_r}_{b_1 \dots b_s} C^{a_1}{}_{cd} - \dots - \alpha^{a_1 \dots a_{r-1} d}_{b_1 \dots b_s} C^{a_r}{}_{cd}$; and t_a is a covector field which may (locally) be defined from t_{ab} via $t_{ab} = t_a t_b$ in a 'temporally orientable' spacetime—for details, see [128, pp. 250-251].

[33]I say 'partly' rather than 'fully' dynamical in light of the compatibility conditions (6.3.2) and (6.3.3), which hold also for $\hat{\nabla}_a$.

and one also eliminates the gravitational potential φ from KPMs of the theory—so that they are quintuples $\langle M, t_{ab}, h^{ab}, \hat{\nabla}_a, \rho \rangle$. (6.3.9) is the geometrised version of Poisson's equation (6.3.7); (6.3.10) holds in a classical spacetime iff this admits, at least locally, a smooth, unit timelike field ξ^a that is geodesic ($\xi^b \nabla_b \xi^a = 0$) and twist-free ($\nabla^{[a}\xi^{b]} = 0$) [128, p. 281]; (6.3.11) holds throughout M iff parallel transport of spacelike vectors in M is, at least locally, path-independent [128, p. 279]. The resulting theory is known as 'Newton-Cartan theory' (NCT).

It can be shown—via Trautman's geometrisation and recovery theorems[34]—that the class of models of NGT which differ by a dynamic shift all map (up to isomorphism) to the same model of NCT. Moreover, for all timelike curves of M with four-velocity field ξ^a, particles subject to a gravitational force in NGT (so that $\xi^b \nabla_b \xi^a = -\nabla^a \varphi$) move along geodesics in NCT (so that $\xi^b \hat{\nabla}_b \xi^a = 0$). In other words, in NCT gravity is no longer a force, as in NGT.

For my purposes, the crucial point to note is the following: by reformulating NGT *à la* NCT as per the above, one constructs a mathematical reformulation of the theory which eliminates the gauge redundancy (in the sense of §6.2) manifest in the possibility of a dynamic shift in NGT. Note also the important point that moving to NCT is not by itself sufficient to be able to understand as physically equivalent all symmetry-related models of Newtonian theory set in flat spacetime. This is because—as mentioned in the previous paragraph—such symmetry-related models will typically correspond to a single model of NCT only up to isomorphism. Thus, in order to have a fully transparent understanding of how it is that symmetry-related models of Newtonian theory set in flat spacetime can correspond to a single model of NCT, a sophisticated substantivalist conception of spacetime ontology is also required.[35]

[34] For original sources, see [209]; for contemporary discussion and proofs, see [128, pp. 267ff.].

[35] A similar moral applies in the case of moving to Galilean spacetime as a response to NGT's boost invariance. Many thanks to David Wallace for this point.

To summarise: According to the interpretational approach, it is typically legitimate *ab initio* to regard symmetry-related models as being physically equivalent, even absent a coherent explication of their common ontology. According to the motivational approach, by contrast, it is not legitimate *ab initio* to regard symmetry-related models as being physically equivalent. Rather, symmetries invariably at most *motivate one to seek* a coherent explication of the common ontology underpinning such models' physical equivalence. When the symmetry-related models in question are not isomorphic, one is motivated to mathematically reformulate the theory. On the other hand, when they are isomorphic, one is not motivated to mathematically reformulate the theory: adopting 'moderate structuralism'—which, in the spacetime context, I take to be equivalent to sophisticated substantivalism[36]—is invariably sufficient.

6.3.3 Motivating Motivation

Up to this point, I have remained officially neutral between the interpretational and motivational approaches to symmetry transformations. Here, however, developing upon [142, §4], I wish to argue explicitly for the latter. One argument in favour of this position is the following: even if the central claim of the interpretational approach—that one may legitimately regard certain symmetry-related models of a theory as being physically equivalent even in the absence of a coherent picture of their common ontology—is true, on this approach, the *reality in terms of which* this physical equivalence is to be understood will, absent further details, remain opaque. That is, without further work, the advocate of the interpretational approach offers no explanation as

[36] I draw the term 'moderate structuralism' from [69]; compare also the 'modest structuralism' of [168, p. 102]. According to this view, objects (e.g. points of spacetime) are construed as being nothing more (or less) than 'nodes' in the relational structures in which they are embedded; and the possibility of purely haecceitistic distinctions between worlds is denied. Construed in this way, moderate structuralism encompasses sophisticated substantivalism—but is a stronger thesis due to the latter clause.

to how such physical equivalence is to be construed, or how it could even be said to arise. To the extent that the interpretational view is not supposed to reduce to an uninteresting form of instrumentalism, it is unclear what realistic picture of the world is being propounded by the defender of this position; it is opaque what, according to her, *the world really is like*.

Here, one must separate two closely-related points. First, the advocate of the interpretational approach may be regarded as shirking her responsibility to provide a coherent explication of the common ontology associated with symmetry-related models; this, however, is where much of the most interesting work in the foundations of physics is done.[37] Second, the advocate of the interpretational approach makes at the outset an assumption that certain symmetry-related models admit of a coherent interpretation which makes manifest their physical equivalence.[38] In my view, it is more cautious to avoid such an assumption: to only regard symmetry-related models as being physically equivalent once an explication of their common ontology can be provided; to consider us always motivated to attempt to construct such an explication; and thereby to favour the motivational over the interpretational approach.

Having said this, it is worth distinguishing two sub-views within the motivational approach. According to the former, more *confident* view, symmetry-related models may only be regarded as being physically equivalent once an interpretation affording a coherent explication of their common ontology is provided, *but such an interpretation is always guaranteed to exist*. By contrast, according to the latter, more *cautious*

[37]For example, the staunchest advocate of the interpretational approach would likely not be motivated to consider whether NGT can be reformulated in terms of Galilean spacetime, or NCT: the bare assertion that models of NGT related by kinematic or dynamic shifts are physically equivalent effectively eliminates motivation for the advocate of the interpretational approach to pursue this research programme.

[38]Cf. footnote 8. Here, the advocate of the interpretational approach may be guided by overarching, *a priori* principles, connecting certain features of the symmetry-related models under consideration with their physical equivalence. However, unless a necessary connection between such features and the physical equivalence of the models can be forged, the point in the body of this paragraph stands.

view, symmetry-related models may only be regarded as being physically equivalent once a coherent explication of their common ontology is provided, *and there is no guarantee that such an interpretation exists*. It should be clear from the foregoing that I favour the latter, more cautious strand of motivationalism.[39] Articulating these two distinct versions of the motivational approach will prove illuminating, when it comes to constructing a taxonomy of the views of philosophical authors in the parallel case of dualities—cf. in particular §6.5.3.

6.4 Equivalence and Duality

In the previous section, I distinguished the interpretational and motivational approaches to symmetry transformations, and defended the latter, expanding upon [142, §4]. In this section, I introduce the notions of 'theoretical equivalence' and 'underdetermination', both of which will prove important in my defence of the motivational approach to dualities in §6.5.

6.4.1 Theoretical Equivalence and Duality

I now introduce a notion of 'theoretical equivalence'. Given two theories $\mathcal{T}_1$ and $\mathcal{T}_2$, with respective spaces of DPMs $\mathcal{D}_1 \subset \mathcal{K}_1$ and $\mathcal{D}_2 \subset \mathcal{K}_2$, I say (broadly following [223, 225, 229]) that these are 'theoretically equivalent' iff (i) there exists an isomorphism between $\tilde{\mathcal{D}}_1$ of $\mathcal{T}_1$ and $\tilde{\mathcal{D}}_2$ of $\mathcal{T}_2$;[40] and (ii) the empirical predictions corre-

[39] I owe the nomenclature of 'confident' versus 'cautious' versions of the motivational approach to Jeremy Butterfield.

[40] Recall from §6.2 that, for a given theory $\mathcal{T}$, $\tilde{\mathcal{D}}$ denotes the gauge-reduced space of DPMs of $\mathcal{T}$. Note also that one may introduce a graded notion of theoretical equivalence, by imposing restrictions on the structure of the models preserved by this isomorphism. Though important to note, this latter point will be set aside in this chapter.

sponding to each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ are identical to the empirical predictions corresponding to the associated $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$.[41] If (i) holds of two theories but not (ii), then I say that they are (merely) 'formally equivalent'; if (ii) holds of two theories but not (i), then I say that they are (merely) 'empirically equivalent'.[42] Two theories are theoretically equivalent iff they are formally equivalent and empirically equivalent.

Turn now to the notion of duality: a pervasive phenomenon in string theory, already introduced and considered at length in Ch. 5. In this chapter, I ask the following question: in addition to being theoretically equivalent (indeed, I take in this thesis theoretical equivalence to be the *definition* of a duality), should duality-related models also be understood as being physically equivalent—i.e., as representing the same possible world? Before I attempt to answer this question, two related points are worth stating. First, dualities are (more) analogous to the kinematic and dynamic shifts than to the static shifts of §6.3.2.4—for the reason that dual models are generically not isomorphic: straightforwardly understood, they represent worlds which differ more than purely with regard to which particular objects are playing which qualitative roles (see e.g. [176, p. 224]). Hence—as I will elaborate—dualities in general motivate mathematical reformulation; adopting moderate structuralism is by itself insufficient to understand satisfactorily the (alleged) physical equivalence of duality-related models. Second, the apparent physical difference between duality-related models can be much more striking than in the case of e.g. models of NGT related by kinematic and dynamic shifts: naïvely understood, the possible worlds they represent are *very* different. For instance, models related by an AdS/CFT-type duality differ in the number of dimensions they (appear to) attribute to spacetime; models related by mirror symmetry differ in the topology they (appear to) attribute to spacetime.

[41] Each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ which correspond to the same empirical predictions in this way may be said to be 'empirically equivalent'—cf. §6.3.1. Note that two models may be empirically equivalent without the theories to which they belong being empirically equivalent, in the sense given below.

[42] The map between $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ may not be one-one in the absence of formal equivalence.

6.4.2 Underdetermination

The distinction between formal and empirical equivalence is of value when discussing whether a pair of theories exhibits 'strong underdetermination of theory by evidence'. I say that two theories $\mathcal{T}_1$ and $\mathcal{T}_2$ present such a case when they are empirically equivalent, yet there exists at least one pair of empirically equivalent models $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ which are nevertheless interpreted as corresponding to distinct possible worlds, respectively W_1 and W_2.

To have a case of strong underdetermination, the two theories must be empirically equivalent; however, one may ask whether formal equivalence is also relevant. Distinguish:

(A) $\mathcal{T}_1$ and $\mathcal{T}_2$ being empirically equivalent *and* formally equivalent.

(B) $\mathcal{T}_1$ and $\mathcal{T}_2$ being empirically equivalent *but not* formally equivalent.

Any argument to the effect that instances of (A) cannot lead to strong underdetermination[43] is (roughly) in line with the Quinean position according to which theories related by reconstrual of predicates (the analogue of formal equivalence) are understood not to lead to such underdetermination [174].[44,45] However, consider again models related by e.g. the AdS/CFT correspondence, or mirror symmetry. In spite of

[43] I.e., to the effect that each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and its associated $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ must correspond to the same world.

[44] For a critical discussion of the Quinean approach to theoretical equivalence, see [10]. With the conclusion of that paper—"If one takes Quine equivalence as the standard for theoretical equivalence, one underestimates the threat of underdetermination" [10, p. 483]—I am in agreement. Cf. also [11].

[45] To allay any possible misunderstanding: Quine's view on this matter is *not* that theories with isomorphic spaces of solutions are always theoretically, or even empirically, equivalent. Rather, Quine holds a strictly stronger view on what it is for two theories to be theoretically equivalent: *if* theories are related by a suitable reconstrual of predicates, *then* such theories are theoretically equivalent. This, plausibly, *entails* that their respective spaces of solutions are isomorphic. But, for Quine, the fact that theories have isomorphic spaces of solutions does not by itself entail that they are theoretically or even empirically equivalent.

an instantiation of formal equivalence, such models of these theories at least *appear* to be ontologically distinct, and thus to correspond to distinct possible worlds.[46] Now, when it comes to deciding whether (A) can indeed lead to strong underdetermination, one can be aided by one's prior commitments in the philosophy of science: even if one is a realist, it may be that by e.g. giving some structuralist account, one can make plausible that models of such theories correspond to the same world (in this regard in the context of dualities, see [131]).[47,48] Nonetheless, examples such as this demonstrate that instances of (A) might, *prima facie*, give rise to strong underdetermination (cf. [131, p. 474])—*pace* Quine, who argued in [174] that only instances of (B) could constitute genuine cases of strong underdetermination.[49] The relevance of this for my views on the interpretation of dualities will become apparent in §6.5.

[46]The concern, therefore, is over the adequacy of formal equivalence—which is, indeed, a formal notion—to capture an informal or semantic notion: that of two models representing the same possible world. My thanks to Jeremy Butterfield for suggesting that I put the matter in this way.

[47]This said, Rickles has recently suggested that cases such as the AdS/CFT correspondence may give rise to *structural* underdetermination [184, 185, 187]. (I concur with this view; cf. footnote 48 below.) Note that if this is so, then even the structuralist may not be able to argue that such pairs of formally equivalent models correspond to the same possible world.

[48]In my view, adopting structural realism as a means of identifying symmetry-related models succeeds only if the models in question are 'naïvely' understood as representing at most haecceitistically distinct possible worlds. In that case, it is clear how adopting structural realism allows us to identify such (putatively) distinct physically possibilities as (actually) not distinct after all. However, if the models in question are 'naïvely' understood as representing more than haecceitistically distinct possible worlds, then adopting structural realism (by itself) is insufficient to provide grounds for understanding the models in question as corresponding to the same possible world.

[49]If this is correct, then I concur with De Haro *et al.* that "we will need to allow that formal isomorphisms do not in general imply sameness of content" [53, §3.1]. Cf. footnote 46.

6.5 Duality as Motivation

6.5.1 Interpretation and Motivation, Reprise

Above, we saw the *prima facie* plausibility of interpreting duality-related models as representing distinct possible worlds. This, however, runs against a common view, that duality-related models may typically be taken *ab initio* to represent the same possible world. For example, Rickles writes:[50]

> [D]ual theories are simply examples of theoretically equivalent descriptions of the same underlying physical content: I distinguish them from cases of genuine underdetermination on the grounds that there is no real incompatibility involved between the descriptions. The incompatibility is at the level of purely unphysical structure. I argue that dual pairs are in fact very strongly analogous to gauge-related solutions ... I conjecture that dualities always point to a more fundamental (intrinsic) description, namely that in which the representational redundancy is eliminated. [187, p. 62]

This position bears striking similarity to the interpretational approach to symmetries.[51] Indeed, by analogy, one may define at this juncture an interpretational approach to dualities. According to this, when presented with a pair of duality-related theories, we are (a) typically warranted in first interpreting duality-related models as representing the same possible world—even absent a coherent explication of their

[50] For further clear expression of this position, see e.g. [133, 134].

[51] One might argue that the final sentence here is in line with the motivational approach. Even if this is true, however, in my view it is not correct to read Rickles as *endorsing* the motivational approach, in light of the preceding sentences in the quote—see below.

common ontology;[52] then may (but are not required to) go on to (b) identify those pairs of dual models, thereby constructing the space of KPMs $\bar{\mathcal{K}}$ of a new theory $\bar{\mathcal{T}}$, which represents the 'common core' (in the language of [48, §2.2], and Ch. 5) of the duality-related theories; and (c) having constructed such a $\bar{\mathcal{K}}$, seek to provide a coherent picture of the ontology of the models $\bar{\mathcal{M}} \in \bar{\mathcal{K}}$.

In contrast with this interpretational approach to dualities, one may also define a motivational approach to dualities. According to this view, the existence of a duality between two theories first (a) motivates us to provide a coherent picture of the common ontology of the pairs of models of these two theories related by the duality; but only once such a characterisation is constructed should we (b) interpret those models as representing the same possible world; and (potentially) (c) identity those models to construct a space of KPMs $\bar{\mathcal{K}}$ of some new theory $\bar{\mathcal{T}}$, which represents the 'common core' of the duality-related theories.

It is important to distinguish two sub-positions within the interpretational view. On the first such view, the existence of a duality motivates us to seek a clear picture of the ontology alleged to underlie the dual-related models.[53] On the second view—by contrast—no such search for a coherent understanding of the common ontology of the dual models is required.[54] Importantly, however (and to reiterate), both of these sub-positions are consistent with the interpretational view. By contrast, the advocate of the motivational approach to dualities maintains that it is *only* legitimate to regard duality-related models as being physically equivalent *if* one possesses a clear picture of the common ontology of the dual models.

Following §6.3.3, one can also introduce two sub-positions within the motiva-

[52]Here, the same points from §6.3.1 regarding the 'typicality' clause arise again.

[53]This view appears more popular in the literature; cf. again e.g. [187, p. 62], and footnote 51.

[54]This can be considered the analogy of Dewar's approach to symmetries in the case of dualities [54, p. 322].

tional approach to dualities—according to the former, more *confident* view, duality-related models should not be regarded as being physically equivalent absent a coherent explication of their common ontology, but such an explication is always guaranteed to be found; according to the latter, more *cautious* view, duality-related models should not be regarded as being physically equivalent absent a coherent explication of their common ontology, and no such explication is guaranteed to be found. This distinction between sub-positions within the motivational approach will prove to be illuminating in §6.5.3, when I consider the views of certain authors *vis-à-vis* the interpretation of duality-related models.

Regardless of where they may stand in the above debate, many authors—both physicists and philosophers—maintain that string-theoretic dualities motivate us to seek a mathematical reformulation of the dual string theories.[55] Since duality-related models are generically non-isomorphic, I concur with this verdict (cf. §6.4.1). Where the advocate of the motivational approach disagrees with some of such authors, however, is on the question of whether, *in the absence of any mathematical reformulation of string theory*, it is legitimate to regard duality-related models as being physically equivalent: in her view, it is not. Moreover, it is precisely on this issue that I disagree with the commonly-held interpretational view.[56]

[55] Or quantum field theories, in the case of e.g. the AdS/CFT correspondence.

[56] Perhaps it is true that many recent philosophical authors' views on dualities are more subtle than a straightforward endorsement of the interpretational approach, *à la* Rickles [187, p. 62]. This notwithstanding, however, a reader unfamiliar with the literature on dualities may obtain the *impression* that the interpretational approach is widely embraced. Here is some *prima facie* evidence to support this claim: "[D]ual [theories] should be understood as giving physically equivalent descriptions" [100, pp. 87-88]; "In all dualities, it is the theories that are equivalent. ... [C]ertain transformations 'don't matter'. The only difference between these [dual] transformations and standard gauge symmetries is that they seem to relate things that *look like they really should matter*!" [187, p. 64]; "[Our] conception of duality meshes with two dual theories being 'gauge-related', in the general philosophical sense of being physically equivalent. For a string duality, such as T-duality and gauge/gravity duality, this means taking such features as the radius of a compact dimension, and the dimensionality of spacetime, to be 'gauge"' [53, p. 68]; "The stance adopted [in this paper] is ... to avoid a literal reading of the elementary/composite interchange and, on this basis, to avoid mixing the question of its meaning with the question of physical fundamentality. The attitude is analogous to the one shared in this volume [a recent special issue of *Studies in the History and Philosophy of Modern Physics* devoted to dualities] about how to understand apparently puzzling features such as the interchange of tiny and huge

Given the importance of this point for the purposes of this chapter, it is worth repeating. When presented with two symmetry- or duality-related models, the interpretationalist will typically say that it is legitimate to regard such models as physically equivalent. The motivationalist will deny this: for her, the mere fact that models are related by a symmetry or duality transformation is not a sufficient reason to regard them as physically equivalent. *This* is what crucially separates the interpretationalist and motivationalist positions.

The motivationalist will go on to say that, for any two symmetry- or duality-related models, we are motivated to try to provide an explication of the common ontology that is alleged to underlie them. *The interpretationalist will not always agree*. Some interpretationalists (e.g. Rickles [187]) *will* claim that that we are so motivated. But not all will (e.g. Dewar [54]). In other words, the interpretationalist and the motivationalist *do not invariably agree about motivation*. Thus, *merely* claiming that dualities motivate us to formulate (e.g.) 'M-theory'[57]—which is conjectured to be the theory which would transparently explain dual string theories' physical equivalence—is not sufficient to make one a motivationalist.

To close this subsection, it is worth reflecting further on the nature of the ontology 'common' to two dual models; two broad attitudes are possible here. First, given two theories understood to be dual, one may attempt to identify the 'shared structure' across duality-related models;[58] one may then use this as a guide to the interpretation of the dual theories (cf. Ch. 5). This austere approach to the interpretation of dualities is advanced in e.g. [48, §2.2]. On the other hand, one may be more ambi-

dimensions connected with T-duality in string theory, or the duality of dimension under the AdS/CFT (gauge/gravity) correspondence. The underlying idea is that, what the dual descriptions do not agree upon, should not be attributed a real physical significance. In fact, this means nothing else than saying that the physics (including its ontology) remains the same under the duality. What changes, is just the way of looking at it" [37, p. 101].

[57] See below for further discussion of this theory. Cf. also Ch. 5.

[58] That is, the formal structure preserved across duality-related models—cf. [48,55]. This is the 'common core' of the two dual models, in the sense given above.

tious. For example, one may construct a new theory, such that dual models of the original theories each constitute (partial) descriptions of certain models of the new theory. The approach of attempting to find an overarching 'M-theory', of which all five superstring theories are 'limits' (in some appropriate sense—cf. §6.5.2), fits naturally into this latter category (again, cf. Ch. 5). It is worth remarking, however, that as it stands the existence of such a theory—as well as our ability to discover it—remains conjectural; for further discussion, see [184, §5.2] and [176, §5].[59]

6.5.2 Motivating Dualities as Motivation

As with the debate between advocates of the interpretational and motivational approaches to symmetries (cf. §6.3.3), I endorse the latter approach to dualities over the former. My reasons for doing so broadly mirror those given in §6.3.3. First, just as in the case of symmetry transformations, the interpretationalist may be regarded as shirking her responsibility to provide a coherent explication of the common ontology associated to duality-related models. Second, the interpretationalist assumes at the outset that certain duality-related models admit of a coherent interpretation which makes manifest their physical equivalence. In my view, however, it is more cautious to drop such an assumption: to only regard duality-related models as being physically equivalent once an explication of their common ontology can be provided; to consider us always motivated to attempt to construct such an explication; and thereby to favour the motivational over the interpretational approach to dualities. (I return in a moment to the distinction between two strands of motivationalism drawn in §§6.3.3 and 6.5.1.)

With the above in mind, recall now that some authors, such as Polchinski, *de-*

[59]For a recent, detailed clarification of the distinction between these two approaches to explicating the ontology 'common' to duality-related models, see [50].

fine duality such that "we have a single quantum system that has two classical limits" [166, p. 7]—from which one concludes that "it is fruitless to argue whether [$\mathcal{T}$] or [$\mathcal{T}'$] provide [*sic*] the fundamental description of the world; rather, it is the full quantum theory" [166, p. 7].[60] If one approaches dualities in this manner, then one begins with a single theory—models of which may be interpreted as corresponding to certain possible worlds—and constructs two further (dual) theories therefrom. In that case, one is already in possession of a coherent account of the physical equivalence of models of the two dual theories, in terms of the ontology of models of the 'quantum' theory from which they are constructed.

If one follows this approach, then one may argue that the interpretational account of dualities is favoured over the motivational—for one may indeed interpret duality-related models as being physically equivalent *ab initio*, via an explication of the ontology of their underlying 'quantum' theory. Note, however, that this is also consistent with the motivational account: since such an explication can be provided in every case, motivation for regarding the duality-related models as being physically equivalent is automatically secured—so we are indeed warranted in doing so.

One may, however, question whether Polchinski's account of dualities is most appropriate—for it is not the case that we always construct dual theories from such an underlying theory. For example, in the case of the AdS/CFT correspondence, neither the string theory in anti-de Sitter space, nor the boundary conformal field theory, was constructed from a third, 'quantum' theory. Indeed, this is also true in the case of string-theoretic dualities: though certain dual string theories are conjectured to be limits of an underlying so-called M-theory (introduced by Witten in [233]), the

[60] Such an approach to dualities is also implicit in [133, 134]. In these passages, by a "quantum system", Polchinski means a quantum theory; by two "classical limits", he means two theories for which perturbation theory is applicable, which may be defined from the original quantum theory—see [166, pp. 6ff.].

existence of this theory was postulated *post facto*[61]—and it is not the case that these string theories were initially defined therefrom. Nevertheless, there is overwhelming evidence (in terms of matching of correlation functions, etc.; see e.g. [14,52]) that such theories are dual, in the sense of §6.4.1.[62]

It is these cases—in which a duality between theories is discovered after the fact, and the existence of an underlying, third theory is only conjectured—which are most interesting from the point of view of the interpretational/motivational distinction. In such cases, I take it that one needs to give a coherent account of the shared ontology (i.e. an appropriate interpretation) of duality-related models, before one declares those models to be physically equivalent; moreover, there appears to exist no set of *a priori* principles by which one may deductively infer that such an interpretation exists. For these reasons, I believe that (a) the definition of duality presented in this chapter is broader, more flexible, and more faithful to string theory history than that proposed by Polchinski; and (b) that the right approach to such dualities is the motivational approach.[63]

6.5.3 Confident and Cautious Motivational Approaches

Other authors working in the foundations of quantum gravity—most notably De Haro—also appear to endorse something akin to the motivational approach to dualities:

> Duality in mathematics is a formal phenomenon: it does not deal with

[61]The *facto* here being the construction of the two original, dual theories.

[62]Specifically, *vis-à-vis* both their theoretical equivalence and their empirical equivalence.

[63]If one endorses these views, then one will argue that more needs to be done to demonstrate the physical equivalence of duality-related models of certain string theories—in terms of exploring the mathematics and interpretation of M-theory—before such physical equivalence can be declared by appeal to this (conjectured) theory.

> physically interpreted structures ... But this is also how the term is used by physicists: it is attached to the equivalence of the formal structures of the theories, regardless of their interpretations, i.e. without it necessarily implying the physical equivalence of the theories which describe two concrete systems. ...
>
> Duality, then, is one of the ways in which two theories can be theoretically equivalent, without its automatically implying their physical equivalence. [48, p. 9]

I concur with this verdict, though it is worth exploring further De Haro's position. To do so, first follow De Haro in distinguishing (i) 'extendable' and 'unextendable' theories—i.e. "theories which do, respectively do not, admit suitable extensions in their domains of applicability" [48, p. 4]—and (ii) 'external' and 'internal' interpretations of theories—i.e. "interpretations which are obtained from outside (i.e. by coupling the theory to a second theory which has already been interpreted), respectively from inside, the theory" [48, p. 4].[64]

De Haro argues that, since extendable theories may be coupled to further theories, they should *not* be regarded as being physically equivalent.[65] His reasoning here is motivated by Galileo ship-type scenarios. To see this, consider the same physical system (e.g. the ship), coupled to two further *different* physical systems (e.g. the shore at rest versus the shore in motion)—we should (De Haro claims) not necessarily regard these two extendable models as being physically equivalent *tout court*, in light of possible couplings to other physical systems which suffice to reveal their physical distinctness. The same point applies to *dual* extendable theories: though they are intertranslatable via the formal duality map, they should *not* be regarded as being

[64] For further details, see [48, §1]. Cf. also [55], in which a very similar distinction is drawn.

[65] For further discussion of these matters in the symmetries literature, see [89, 205]; cf. also footnote 10.

physically equivalent, in light of possible further couplings to other physical systems.

Though I concur with De Haro on this point—indeed, ultimately I *fully* agree with De Haro (and his close collaborator, Jeremy Butterfield) on these matters—it is worth reflecting further upon his writings (and those of Butterfield), since doing so brings to light two different sub-views within the motivational approach (*viz.*, the *confident* and *cautious* versions of the motivational approach, discussed in §§6.3.3 and 6.5.1). To this end, consider the following passage from Butterfield: [32, p. 5]

> ... De Haro proposes a sufficient condition for one to be justified in interpreting two duals to be physically equivalent. This sufficient condition has two conjuncts. The first is that each dual is *unextendable*: which means, roughly speaking, that the dual, i.e. the theory, both is a complete description of its intended domain and cannot be extended to a larger domain. The second is that each dual has an *internal* (as against *external*) interpretation: i.e. an interpretation that does *not* proceed by coupling to another theory, often one which describes measurements of the given theory's domain. These conjuncts are linked in that De Haro argues that unextendability implies that one is justified in using an internal interpretation: (justified but not obliged—there can still be external interpretations).

I fully agree with this passage. Indeed, Butterfield goes on to write that "a statement of the bare theory, and an internal interpretation, are not automatic, given a proven duality" [32, p. 43], and to argue that (e.g.) Newton and Clarke were right not to move from Newtonian mechanics formulated in Newtonian spacetime to Newtonian mechanics formulated in Galilean (or Newton-Cartan) spacetime, as they did not have the relevant internal interpretation to hand (cf. [45, p. 854]). This resistance to *ab initio* declarations of physical equivalence is, of course, very consonant with the motivational approach—to which, for such reasons, I take Butterfield and De Haro to subscribe.

That said, the above, extended passage from Butterfield is ambiguous between a 'confident' version of motivationalism, according to which the internal interpretation on which dual (unextendable) theories may be regarded as being physically equivalent always exists, and a 'cautious' version of motivationalism, according to which there is no guarantee that such an internal interpretation exists. One of the central contributions of the present chapter is the delineation of these distinct sub-views, hitherto overlooked in the literature.[66]

As I see it, the cautious version of the motivational approach is to be preferred. Let me explain why; I have, in particular, three points to make. First, it seems possible to envisage (admittedly artificial) cases in which duality-related models do *not* possess sufficiently rich common structure to afford a coherent physical interpretation: imagine that the two models agree on empirical substructures, in addition to having extra structure, which is not isomorphic between the two models. In that case, the only 'common' structures may be the empirical substructures of both models—but the empirical substructures alone are insufficient for a *realist* understanding of what the mathematical model is supposed to represent in the world.

Second, some philosophers, notably Dasgupta, would maintain that we do *not* yet have a coherent explication of the common ontology underlying models of Newtonian mechanics related by static shifts, or models of GR related by hole diffeomorphisms—for, Dasgupta argues, standard responses such as sophisticated substantivalism make appeal to problematic "bare modal claims" [43, p. 120]; moreover, any alternative approach, such as an appeal to so-called 'Einstein algebras' (cf. e.g. [64, ch. 9]), will face

[66] I thank Jeremy Butterfield and Sebastian De Haro for discussion on these matters; in private communication, both have indicated that they favour the cautious motivational approach.

similar problems. In such a case, for authors such as Dasgupta (who, incidentally, also endorses something like the motivational approach to symmetries—cf. [45, pp. 853-854]), no such coherent explication of the common ontology underlying the models in question is forthcoming; nor is it guaranteed to exist. Thus, this particular case illustrates one of the ways in which one would be pushed towards the more cautious motivational view.

Third and finally, my preferred version of the motivational approach is, I contend, the most epistemically cautious interpretative attitude possible towards models of physical theories related by symmetries/dualities.[67] In my view, it is therefore unreasonable to demand that the burden of justification lies with this position; rather, there exists a positive burden of justification—here to argue that there always exists a coherent explication of the common ontology underlying symmetry- or duality-related models—for advocates of riskier approaches, such as the confident motivational approach, or the interpretational approach.

6.6 Consensus

In §6.1, I highlighted two oft-advanced reasons for regarding duality-related models as being physically equivalent in the absence of a coherent explication of their common ontology: (1) such a position—the interpretational approach to dualities—fits with the parallel interpretational approach to symmetries; (2) such a position accords with a perceived consensus in contemporary theoretical physics. Up to this point, I have focussed on arguing against (1)—by arguing against the interpretational ap-

[67] There is some analogy here with van Fraassen's constructive empiricism [210], often justified on the grounds that "belief in the empirical adequacy of accepted theories [is] the weakest attitude one can attribute to scientists at the same time that one is still able to make sense of their scientific activity" [144, §2.2].

proaches to both symmetries and dualities, and for their motivational alternatives. In this section, I turn to (2), giving evidence that the interpretational approach to dualities does not, in fact, represent a consensus in the relevant areas of theoretical physics.

To begin, recall from §6.5.3 De Haro's observation that, in physics, the term 'duality' "is attached to the equivalence of the formal structures of the theories, regardless of their interpretations" [48, p. 12]. That is, in theoretical physics the term 'duality' is often applied to theories which are solely theoretically equivalent—physical equivalence notwithstanding. It is not hard to identify explicit evidence for De Haro's claim in the literature. For example, Vafa defines a duality thus: [213, pp. 4-5][68]

> Consider a physical system Q (which I will not attempt to define). And suppose this system depends on a number of parameters $[\lambda_i]$. Collectively we denote the space of the parameters λ_i by $\mathcal{M}$ which is usually called the moduli space of coupling constants of the theory. ... Typically physical systems have many observables which we could measure. Let us denote the observables by $\mathcal{O}_\alpha$. Then we would be interested in their correlation functions ... The totality of such observables and their correlation functions determine a physical system. Two physical systems $Q\left[\mathcal{M}, \mathcal{O}_\alpha\right]$, $\tilde{Q}\left[\tilde{\mathcal{M}}, \tilde{\mathcal{O}}_\alpha\right]$ are dual to one another if there is an isomorphism between $\mathcal{M}$ and $\tilde{\mathcal{M}}$ and $\mathcal{O} \leftrightarrow \tilde{\mathcal{O}}$ respecting all the correlation functions.[69]

Vafa's focus is clearly upon the formal, mathematical equivalence of the two theories in question (i.e. upon formal equivalence, in the sense of §6.4.1). If one understands correlation functions as being the empirical substructures of the theories under

[68] For philosophical discussions drawing upon Vafa's definition of a duality, see [134, §3] and [176, §2].

[69] Here, by a 'physical system' Q, we can understand Vafa to mean a physical theory $\mathcal{T}$; one may then identify $\mathcal{M}$ with our reduced space of DPMs, $\tilde{\mathcal{D}}$ (note that the tilde here refers to the quotienting of $\mathcal{D}$ by gauge-equivalent models, rather than to the space of DPMs of the dual theory). By $\mathcal{O}$, we understand Vafa to mean the set of observables $\mathcal{O}_\alpha$ for the theory in question.

consideration (in the sense of [210, pp. 67ff.]), then one may be able to argue in addition that Vafa is interested in the empirical equivalence of such theories. However, for my purposes the essential point here is that no claim regarding the physical equivalence of duality-related models is advanced. This accords with the view that, in the physics literature, the term 'duality' is often applied to cases of theoretical equivalence, without (explicit) assumptions being made regarding the physical equivalence of the models in question.

For a second piece of evidence to this end, consider the following quote from Maldacena [130, p. 61], made in the context of the AdS/CFT correspondence:

> What does it really mean for the two [dual] theories to be equivalent? First, for every entity in one theory, the other theory has a counterpart. The entities may be very different in how they are described by the theories: one entity in the interior might be a single particle of some type, corresponding on the boundary to a whole collection of particles of another type, considered as one entity. Second, the predictions for corresponding entities must be identical. Thus, if two particles have a 40 percent chance of colliding in the interior, the two corresponding collections of particles on the boundary should also have a 40 percent chance of colliding.

Maldacena is making two claims here. First, he is saying that, given two dual theories, every entity in each theory has a 'counterpart' entity in the corresponding dual theory. Second, he is saying that dual theories must be empirically equivalent. With regard to the first point: to say that each such entity has a 'counterpart' in the corresponding dual theory is, of course, very different from saying that they are truly one and the same thing. (The relation of counterparthood is not the identity relation!) And with regard to the second point: empirical equivalence is, of course, very different from physical equivalence. (Physical equivalence implies empirical equivalence, but not vice versa.)

To be clear, it is not my intention here to try to answer the (tricky) question of when exactly physicists do and do not call an isomorphism between spaces of solutions a duality. Rather, my purpose in this section is less ambitious: I seek merely to emphasise that physicists do not *invariably* regard dual models as being physically equivalent, and (relatedly) that they do not *invariably* regard the notion of 'duality' as involving the notion of physical equivalence. To do this, it is clearly not required that one gets a grip on the question of when exactly physicists call an isomorphism between spaces of solutions a 'duality'—though I admit that this is indeed an interesting question worthy of further scrutiny.

6.7 Close

In this chapter, I have argued that it is not invariably legitimate to regard duality-related models as being physically equivalent; rather, the existence of a duality at most *motivates one to seek* a coherent explication of the ontology underpinning their physical equivalence—and only once such an explication is secured may one indeed take the models in question to be physically equivalent. In order to achieve this goal, I have both (in §6.5) appealed to an analogous distinction in the case of dualities to that between the interpretational and motivational approaches to symmetry transformations; and argued (in §6.6)—for what it is worth—that physicists' understanding of duality-related models *vis-à-vis* their physical equivalence is less clear-cut than some philosophers take it to be. The moral of these investigations is a familiar one in the philosophy of physics: that the situation—here regarding dualities—is less straightforward than one might at first think. This chapter may be judged a success to the extent that the (in my view) over-simplified, false impression that duality-related models may always be taken to represent the same physical state of affairs is dispelled.

Bibliography

[1] Pablo Acuña, "Minkowski Spacetime and Lorentz Invariance: The Cart and the Horse or Two Sides of a Single Coin?", Studies in History and Philosophy of Modern Physics 55, pp. 1-12, 2016.

[2] Gerardo Aldazabal, Diego Marqués and Carmen Núñez, "Double Field Theory: A Pedagogical Review", Classical and Quantum Gravity 30, 163001, 2013.

[3] H. G. Alexander (ed.), *The Leibniz-Clarke Correspondence: Together with Extracts from Newton's Principia and Opticks*, Manchester: Manchester University Press, 1986.

[4] Martin Ammon and Johanna Erdmenger, *Gauge/Gravity Duality: Foundations and Applications*, Cambridge: Cambridge University Press, 2015.

[5] James L. Anderson, *Principles of Relativity Physics*, New York: Academic Press, 1967.

[6] Felipe A. Asenjo and Sergio A. Hojman, "Do Electromagnetic Waves Always Propagate Along Null Geodesics?", Classical and Quantum Gravity 34, 205011, 2017.

[7] David Baker, "Spacetime Substantivalism and Einstein's Cosmological Constant", Philosophy of Science 72(5), pp. 1299-1311, 2005.

[8] David Baker, "Symmetry and the Metaphysics of Physics", Philosophy Compass 5, pp. 1157-1166, 2010.

[9] Julian Barbour, "Shape Dynamics: An Introduction", in *Proceedings of the Conference Quantum Field Theory and Gravity*, Regensburg, 2010.

[10] Thomas William Barrett and Hans Halvorson, "Glymour and Quine on Theoretical Equivalence", Journal of Philosophical Logic 45(5), pp. 467-483, 2016.

[11] Thomas William Barrett and Hans Halvorson, "Morita Equivalence", The Review of Symbolic Logic 9(3), pp. 556-582, 2016.

[12] Yuri Balashov and Michel Janssen, "Presentism and Relativity", British Journal for the Philosophy of Science 54(2), pp. 327-346, 2003.

[13] Hans Bauer, "Über die Energiekomponenten des Gravitationsfeldes", *Physikalische Zeitschrift* 19, pp. 163-166, 1918.

[14] Katrin Becker, Melanie Becker, and John Schwarz, *String Theory and M-Theory: A Modern Introduction*, Cambridge: Cambridge University Press, 2007.

[15] Jacob D. Bekenstein, "An Alternative to the Dark Matter Paradigm: Relativistic MOND Gravitation", invited talk at the 28th Johns Hopkins Workshop on Current Problems in Particle Theory, June 2004, Johns Hopkins University, Baltimore. Available at arXiv:astro-ph/0412652.

[16] Jacob D. Bekenstein, "Relativistic Gravitation Theory for the MOND Paradigm", available at arXiv:astro-ph/0403694.

[17] Gordon Belot, "Geometry and Motion", British Journal for the Philosophy of Science 51, pp. 561-595, 2000.

[18] Gordon Belot, "The Principle of Sufficient Reason", The Journal of Philosophy 98, pp. 55-74, 2001.

[19] Gordon Belot, “Symmetry and Equivalence”, in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, pp. 318-339, 2013.

[20] Gordon Belot, “Fifty Million Elvis Fans Can’t be Wrong”, forthcoming in Noûs, 2017.

[21] P. G. Bergmann, *Introduction to the Theory of Relativity*, Dover, 1976.

[22] Ralph Blumenhagen, Dieter Lüst, and Stefan Theisen, *Basic Concepts of String Theory*, Springer Series in Theoretical and Mathematical Physics, Berlin: Springer, 2013.

[23] Robert Brandenberger and Cumrum Vafa, “Superstrings in the Early Universe”, Nuclear Physics B 316, pp. 391-410, 1989.

[24] Harvey R. Brown, *Physical Relativity: Space-Time Structure From a Dynamical Perspective*, Oxford: Oxford University Press, 2005.

[25] Harvey R. Brown and Peter Holland, “Simple Applications of Noether’s First Theorem in Quantum Mechanics and Electromagnetism”, American Journal of Physics 72, pp. 34-39, 2004.

[26] Harvey R. Brown and Dennis Lehmkuhl, “Einstein, the Reality of Space, and the Action-Reaction Principle”, in Partha Ghose (ed.), *Einstein, Tagore and the Nature of Reality*, London and New York: Routledge, 2016.

[27] Harvey R. Brown and Oliver Pooley, “The Origins of the Spacetime Metric: Bell’s Lorentzian Pedagogy and its Significance in General Relativity”, in Craig Callender and Nick Huggett (eds.), *Physics Meets Philosophy at the Plank Scale*, Cambridge: Cambridge University Press, 2001.

[28] Harvey R. Brown and Oliver Pooley, “Minkowski Space-Time: A Glorious Non-Entity”, in Dennis Dieks (ed.), *The Ontology of Spacetime*, Elsevier, 2006.

[29] Harvey R. Brown and James Read, “Clarifying Possible Misconceptions in the Foundations of General Relativity”, American Journal of Physics 84(5), pp. 327-334, 2016.

[30] Harvey R. Brown and James Read, “The Dynamical Approach to Spacetime Theories”, in E. Knox and A. Wilson (eds.), *The Routledge Companion to Philosophy of Physics*, London: Routledge, 2018. (Forthcoming.)

[31] Jeremy Butterfield, “Reconsidering Relativistic Causality”, International Studies in the Philosophy of Science 21(3), pp. 295-328, 2007.

[32] Jeremy Butterfield, “On Dualities and Equivalences Between Physical Theories”, in N. Huggett, B. Le Bihan and C. Wüthrich, *Philosophy Beyond Spacetime*, Oxford: Oxford University Press, 2018. (Forthcoming.)

[33] Curtis G. Callan, D. Friedan, E. J. Martinec, and M. J. Perry, “Strings in Background Fields”, Nuclear Physics B 262(4), pp. 593-609, 1985.

[34] Craig Callender, *What Makes Time Special?*, Oxford: Oxford University Press, 2017.

[35] Sean Carroll, George Field and Roman Jackiw, “Limits on a Lorentz- and Parity-Violating Modification of Electrodynamics”, Physical Review D 41(4), 1990.

[36] Sean Carroll and Eugene Lim, “Lorentz-Violating Vector Fields Slow the Universe Down”, Physical Review D 70, 123525, 2004.

[37] Elena Castellani, “Duality and ‘Particle’ Democracy”, Studies in History and Philosophy of Modern Physics 59, pp. 100-108, 2017.

[38] Adam Caulton, “The Role of Symmetry in the Interpretation of Physical Theories”, Studies in History and Philosophy of Modern Physics 52, pp. 153-162, 2015.

[39] A. Coley, “Schiff’s Conjecture on Gravitation, Physical Review Letters 49(12), pp. 853-855, 1982.

[40] Richard Creath, “Logical Empiricism”, in *The Stanford Encyclopedia of Philosophy*, 2017.

[41] Erik Curiel, “A Primer on Energy Conditions”, in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, pp. 43-104, 2017.

[42] Erik Curiel, “On Geometric Objects, the Non-Existence of a Gravitational Stress-Energy Tensor, and the Uniqueness of the Einstein Field Equation”, Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[43] Shamik Dasgupta, “The Bare Necessities”, Philosophical Perspectives 25, pp. 115-160, 2011.

[44] Shamik Dasgupta, “On the Plurality of Grounds”, Philosophers’ Imprint 14(20), 2014.

[45] Shamik Dasgupta, “Symmetry as an Epistemic Notion (Twice Over)”, British Journal for the Philosophy of Science 67(3), pp. 837-878, 2016.

[46] Richard Dawid, *String Theory and the Scientific Method*, Cambridge: Cambridge University Press, 2013.

[47] Talal Debs and Michael Redhead, *Objectivity, Invariance, and Convention*, Cambridge, MA: Harvard University Press, 2007.

[48] Sebastian De Haro, “Spacetime and Physical Equivalence”, in N. Huggett, K. Matsubara and C. Wüthrich, *Beyond Spacetime: The Foundations of Quantum Gravity*, Cambridge: Cambridge University Press, 2018. (Forthcoming.)

[49] Sebastian De Haro, “Dualities and Emergent Gravity: AdS/CFT and Verlinde’s Scheme”, Studies in History and Philosophy of Modern Physics 59, pp. 109-125, 2017.

[50] Sebastian De Haro, “The Heuristic Function of Duality”, Synthese, 2018. (Forthcoming.)

[51] Sebastian De Haro and Jeremy Butterfield, “A Schema for Dualities, Illustrated by Bosonization”, in J. Kouneiher (ed.), *Foundations of Mathematics and Physics One Century After Hilbert*, Berlin: Springer, 2018.

[52] Sebastian De Haro, Daniel R. Mayerson and Jeremy N. Butterfield, “Conceptual Aspects of Gauge/Gravity Duality”, Foundations of Physics 46(11), pp. 1381-1425, 2016.

[53] Sebastian De Haro, Nicholas Teh and Jeremy N. Butterfield, “Comparing Dualities and Gauge Symmetries”, Studies in History and Philosophy of Modern Physics 59, pp. 68-80, 2017.

[54] Neil Dewar, “Symmetries and the Philosophy of Language”, Studies in the History and Philosophy of Modern Physics 52, pp. 317-327, 2015.

[55] Neil Dewar, “Interpretation and Equivalence; or, Equivalence and Interpretation”, in E. Curiel and S. Lutz (eds.), *The Semantics of Theories*, 2017. (Forthcoming.)

[56] Neil Dewar, “Sophistication about Symmetries”, British Journal for Philosophy of Science, 2017. (Forthcoming.)

[57] Neil Dewar, “Maxwell Gravitation”, Philosophy of Science 85, pp. 249-270, 2018.

[58] Neil Dewar and James Owen Weatherall, “On Gravitational Energy in Newtonian Theories”, Foundations of Physics 48, pp. 558-578, 2018.

[59] Dennis Dieks, Jeroen van Dongen and Sebastian de Haro, “Emergence in Holographic Scenarios for Gravity”, Studies in History and Philosophy of Modern Physics 52, pp. 203-216, 2015.

[60] Paul Dirac, *The Principles of Quantum Mechanics*. Oxford: Oxford University Press, 1930.

[61] Anthony Duncan, *The Conceptual Framework of Quantum Field Theory*, Oxford: Oxford University Press, 2017.

[62] Gerald V. Dunne, “Aspects of Chern-Simons Theory”, Les Houches Lectures, 1998. Available at arXiv:hep-th/9902115.

[63] Patrick Dürr, “Fantastic Beasts and Where (Not) to Find Them, Part I: Local Gravitational Energy and Energy Conservation in General Relativity”, 2018.

[64] John Earman, *World Enough and Space-Time*, Cambridge, MA: MIT Press, 1989.

[65] John Earman and John Norton, “What Price Spacetime Substantivalism? The Hole Story”, British Journal for the Philosophy of Science 38(4), pp. 515-525, 1987.

[66] Arthur Eddington, *The Mathematical Theory of Relativity*, Cambridge: Cambridge University Press, 1923.

[67] Jürgen Ehlers and Robert Geroch, “Equation of Motion of Small Bodies in Relativity”, Annals of Physics 309, pp. 232-236, 2004.

[68] Jürgen Ehlers, Felix A. E. Pirani and Alfred Schild, "The Geometry of Free Fall and Light Propagation", in L. O'Reifeartaigh (ed.), *General Relativity: Papers in Honour of J. L. Synge*, Oxford: Clarendon Press, pp. 63-84, 1972.

[69] Michael Esfeld and Vincent Lam, "Moderate Structural Realism About Spacetime", Synthese 160(1), pp. 27-46, 2011.

[70] Albert Einstein, "On the Electrodynamics of Moving Bodies", Annalen der Physik 17, pp. 891-921, 1905.

[71] Albert Einstein, "Die formale Grundlage der allgemeinen Relativitätstheorie", *Königlich Preußische Akademie der Wissenschaften*, Berlin, *Sitzungsberichte*, pp. 769-822, 1914. Available at *The Collected Papers of Albert Einstein*, Martin J. Klein, A. J. Kox and Robert Schulman (eds.), *Volume 6: The Berlin Years: Writings, 1914-1917*, doc. 9, pp. 72ff.

[72] Albert Einstein, "Die Grundlage der allgemeinen Relativitätstheorie", *Annalen der Physik* 49(7), pp. 769-822, 1916. Reprinted as Vol. 6, Doc. 30, CPAE.

[73] Albert Einstein, "Über Gravitationswellen", *Königlich Preußische Akademie der Wissenschaften*, Berlin, *Sitzungsberichte*, pp. 154-167, 1918. Available at *The Collected Papers of Albert Einstein*, Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.), *Volume 7: The Berlin Years: Writings, 1918-1921*, doc. 1, pp. 11ff.

[74] Albert Einstein, "Der Energiesatz in der allgemeinen Relativitätstheorie" and "Nachtrag zur Korrektur", *Königlich Preußische Akademie der Wissenschaften*, Berlin, *Sitzungsberichte*, 1918. Available at *The Collected Papers of Albert Einstein*, Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.), *Volume 7: The Berlin Years: Writings, 1918-1921*, doc. 9, pp. 63ff.

[75] Albert Einstein, “What is the Theory of Relativity?”, London: *The Times*, 1919.

[76] Albert Einstein, “The Fundaments of Theoretical Physics”, in *Ideas and Opinions*, New York: Bonanza, pp. 323-335, 1940.

[77] Albert Einstein, letter to Max von Laue, September 12th, 1950. Einstein Archive, Boston (EA 16-148).

[78] Albert Einstein and Marcel Grossmann, “Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie geggründeten Gravitationstheorie”, *Zeitschrift für Mathematik und Physik* 63, 1914. Available at *The Collected Papers of Albert Einstein*, Martin J. Klein, A. J. Kox and Robert Schulman (eds.), *Volume 6: The Berlin Years: Writings, 1914-1917*, doc. 2, pp. 6ff.

[79] Steven French, “The Interdependence of Structure, Objects and Dependence”, Synthese 175(1), pp. 89-109, 2010.

[80] Steven French, *The Structure of the World: Metaphysics and Representation*, Oxford: Oxford University Press, 2014.

[81] Michael Friedman, *Foundations of Space-Time Theories*, Princeton: Princeton University Press, 1983.

[82] Michael Friedman, *Reconsidering Logical Positivism*, Cambridge: Cambridge University Press, 1999.

[83] Robert Geroch and Pong Soo Jang, “Motion of a Body in General Relativity”, Journal of Mathematical Physics 16(1), pp. 65-67, 1975.

[84] Robert Geroch and James Owen Weatherall, “The Motion of Small Bodies in Space-time”, 2017.

[85] Michel Ghins and Tim Budden, “The Principle of Equivalence”, Studies in History and Philosophy of Modern Physics 32, pp. 33-51, 2001.

[86] Peter Godfrey Smith, *Theory and Reality: An Introduction to the Philosophy of Science*, Chicago, IL: University of Chicago Press, 2013.

[87] Hubert Goenner, "Theories of Gravitation with Nonminimal Coupling of Matter and the Gravitational Field", Foundations of Physics 14(9), pp. 865-881, 1984.

[88] Jaume Gomis and Hirosi Ooguri, "Nonrelativistic Closed String Theory", Journal of Mathematical Physics 42, pp. 3127-3151, 2001.

[89] Hilary Greaves and David Wallace, "Empirical Consequences of Symmetries", British Journal for the Philosophy of Science 65, pp. 59-89, 2014.

[90] Michael B. Green, John H. Schwarz and Edward Witten, *Superstring Theory*, Vol. 1, Cambridge: Cambridge University Press, 1987.

[91] James Hartle, *Gravity: An Introduction to Einstein's General Relativity*, San Francisco: Addison Wesley, 2003.

[92] Richard Healey, *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*, Oxford: Oxford University Press, 2009.

[93] Friedrich Hehl and Yuri Obukhov, "How Does the Electromagnetic Field Couple to Gravity, in Particular to Metric, Nonmetricity, Torsion, and Curvature?", in *Testing Relativistic Gravity in Space: Gyroscopes, Clocks, Interferometers*, Bad Honnef, C. Laemmerzahl et al. (eds.). Springer, Berlin, 2000.

[94] José A. Heras, "Electromagnetism in Euclidean Four Space: A Discussion Between God and the Devil", American Journal of Physics 62, pp. 914-916, 1994.

[95] Carl Hoefer, "Energy Conservation in GTR", Studies in History and Philosophy of Modern Physics 31, pp. 187-199, 2000.

[96] Gary T. Horowitz and Joseph Polchinski, "Gauge/Gravity Duality", in Daniele Oriti (ed.), *Approaches to Quantum Gravity: Toward a New Understanding of Space, Time and Matter*, Cambridge University Press, 2009.

[97] Nick Huggett, *Space from Zeno to Einstein*, Cambridge, MA: MIT Press, 1999.

[98] Nick Huggett, "Essay Review: *Physical Relativity* and *Understanding Space-Time*", Philosophy of Science 76, pp. 404-422, 2009.

[99] Nick Huggett, "The String Theoretic Explanation of General Relativity", 2016.

[100] Nick Huggett, "Target Space $\neq$ Space", Studies in the History and Philosophy of Modern Physics 59, pp. 81-88, 2017.

[101] Nick Huggett and Craig Callender, "Why Quantize Gravity (or Any Other Field for That Matter)?", Philosophy of Science 68, pp. S382-S394, 2001.

[102] Nick Huggett and Tiziana Vistarini, "Deriving General Relativity from String Theory", Philosophy of Science 82(5), pp. 1163-1174, 2015.

[103] Nick Huggett and Christian Wüthrich, "Emergent Spacetime and Empirical (In)Coherence", Studies in History and Philosophy of Modern Physics 44, pp. 276-285, 2013.

[104] Jenann Ismael and Bas C. van Fraassen, "Symmetry as a Guide to Superfluous Theoretical Structure", in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, pp. 371-392, 2003.

[105] Yakov Itin and Friedrich Hehl, "Is the Lorentzian Signature of the Metric of Spacetime Electromagnetic in Origin?", Annals of Physics 312, pp. 60-83, 2004.

[106] Ted Jacobson and David Mattingly, "Gravity with a Dynamical Preferred Frame", Physical Review D 64, 024028, 2001.

[107] Michel Janssen,"COI Stories: Explanation and Evidence in the History of Science", Perspectives on Science 10, pp. 457-522, 2002.

[108] Michel Janssen, "Drawing the Line Between Kinematics and Dynamics in Special Relativity", Studies in History and Philosophy of Modern Physics 40, pp. 26-52, 2009.

[109] Felix Klein, "A Comparative Review of Recent Researches in Geometry", Bulletin of the New York Mathematical Society 2, pp. 215-249, 1982. Translation by M. W. Haskell.

[110] Eleanor Knox, "Newton-Cartan Theory and Teleparallel Gravity: The Force of a Formulation", Studies in the History and Philosophy of Modern Physics 42, pp. 264-275, 2011.

[111] Eleanor Knox, "Effective Spacetime Geometry", Studies in History and Philosophy of Modern Physics 44, pp. 346-356, 2013.

[112] Eleanor Knox, "Newtonian Spacetime Structure In Light of the Equivalence Principle", British Journal for the Philosophy of Science 65(4), pp. 863-880, 2014.

[113] Eleanor Knox, "Physical Relativity from a Functionalist Perspective", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[114] James Ladyman, "What is Structural Realism?", Studies in History and Philosophy of Science 29, pp. 409-424, 1998.

[115] James Ladyman, "On the Identity and Diversity of Objects in a Structure", Proceedings of the Aristotelian Society 81, pp. 23-43, 2007.

[116] James Ladyman, "Structural Realism", in *The Stanford Encyclopedia of Philosophy*, 2014.

[117] James Ladyman and Don Ross (with Don Spurrett and John Collier), *Every Thing Must Go: Metaphysics Naturalised*, Oxford: Oxford University Press, 2007.

[118] Vincent Lam, "Gravitational and Nongravitational Energy: The Need for Background Structures", Philosophy of Science 78, pp. 1012-1024, 2011.

[119] A. R. Lee and T. M. Kalotas, "Lorentz Transformations from the First Postulate", American Journal of Physics 43, pp. 434-437, 1975.

[120] Dennis Lehmkuhl, "Is Spacetime a Gravitational Field?", in Dennis Dieks (ed.), *The Ontology of Spacetime II*, Elsevier, 2008.

[121] Dennis Lehmkuhl, "Mass-Energy-Momentum: Only there Because of Spacetime?", British Journal for the Philosophy of Science 62, 2011.

[122] Dennis Lehmkuhl, "Why Einstein Did Not Believe that General Relativity Geometrizes Gravity", Studies in the History and Philosophy of Modern Physics 46, pp. 316-326, 2014.

[123] Dennis Lehmkuhl, "The Equivalence Principle(s)", forthcoming in E. Knox and A. Wilson (eds.), *The Routledge Companion to the Philosophy of Physics*, London: Routledge, 2018.

[124] Tullio Levi-Civita, "Sulla espressione analitica spettante al tensore gravitazionale nella teoria di Einstein", *Rendiconti Accademia dei Lincei* ser. 5, 26, pp. 381-391, 1917.

[125] J.-M. Lévy-Leblond, "One More Derivation of the Lorentz Transformation", American Journal of Physics 44, pp. 271-279, 1976.

[126] Alan P. Lightman and David L. Lee, "Restricted Proof that the Weak Equivalence Principle Implies the Einstein Equivalence Principle, Physical Review D 8(2), pp. 364-376, 1973.

[127] David Malament, "A Remark About the "Geodesic Principle" in General Relativity", in M. Frappier, D. Brown and R. DiSalle (eds.), *Analysis and Interpretation in the Exact Sciences*, The Western Ontario Series in Philosophy of Science, vol. 78, Dordrecht: Springer, pp. 245-252, 2012.

[128] David Malament, *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*, Chicago: University of Chicago Press, 2012.

[129] Juan Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", Advances in Theoretical and Mathematical Physics 2, pp. 231-252, 1998.

[130] Juan Maldacena, "The Illusion of Gravity", Scientific American, November 2005.

[131] Keizo Matsubara, "Realism, Underdetermination and String Theory Dualities", Synthese, pp. 471-489, 2013.

[132] Keizo Matsubara "Quantum Gravity and the Nature of Space and Time", Philosophy Compass 12 (3), 2017.

[133] Keizo Matsubara and Lars-Göran Johansson, "Spacetime in String Theory: A Conceptual Clarification", Journal for General Philosophy of Science, 2018. (Forthcoming.)

[134] Keizo Matsubara and Chris Smeenk, "Dualities and Effective Ontology", 2016.

[135] Tim Maudlin, "Buckets of Water and Waves of Space", Philosophy of Science 60(2), pp. 183-203, 1993.

[136] Tim Maudlin, "Healey on the Aharonov-Bohm Effect" Philosophy of Science 65, pp. 361-368, 1998.

[137] Tim Maudlin, *Philosophy of Physics: Space and Time*. Princeton, NJ: Princeton University Press, 2012.

[138] Flavio Mercati, *Shape Dynamics: Relativity and Relationalism*, Oxford: Oxford University Press, 2018.

[139] Hermann Minkowski, "Raum und Zeit", Physikalische Zeitschrift 10, pp. 104-111, 1909.

[140] Tushar Menon, Niels Linnemann and James Read, "Clocks and Chronogeometry: Rotating Spacetimes and the Relativistic Null Hypothesis", British Journal for the Philosophy of Science, 2018. (Forthcoming.)

[141] Charles Misner, Kip Thorne, and John Wheeler, *Gravitation*, San Francisco: Freeman & Co., 1973.

[142] Thomas Møller-Nielsen, "Invariance, Interpretation, and Motivation", Philosophy of Science 84, pp. 1253-1264, 2017.

[143] Thomas Møller-Nielsen and James Read, "Redundant Epistemic Symmetries", 2017.

[144] Bradley Monton and Chad Mohler, "Constructive Empiricism", in *The Stanford Encyclopedia of Philosophy*, 2017.

[145] F. A. Muller, "The Equivalence Myth of Quantum Mechanics—Part I", Studies in History and Philosophy of Modern Physics 28(1), pp. 35-61, 1997.

[146] F. A. Muller, "The Equivalence Myth of Quantum Mechanics—Part II", Studies in History and Philosophy of Modern Physics 28(2), pp. 219-247, 1997.

[147] Wayne C. Myrvold, "How Could Relativity be Anything Other Than Physical?", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[148] Horatiu Nastase, "Introduction to Supergravity", arXiv:1112.3502v3, 2012.

[149] Graham Nerlich, *Einstein's Genie: Spacetime out of the Bottle*, Montreal: Minkowski Institute Press, 2013.

[150] James Nguyen, Nicholas Teh and Laura Wells, "Surplus Structure is Not Superfluous", British Journal for the Philosophy of Science, 2017. (Forthcoming.)

[151] Wei-Tou Ni, "Equivalence Principles and Electromagnetism, Physical Review Letters 38(7), pp. 301-304, 1977.

[152] John D. Norton, "What Was Einstein's Principle of Equivalence?", Studies in History and Philosophy of Science 16, pp. 203-246, 1985.

[153] John D. Norton, "Why Constructive Relativity Fails", British Journal for the Philosophy of Science 59, pp. 821-834, 2008.

[154] Robert Nozick, *Invariances: The Structure of the Objective World*, Cambridge, MA: Harvard University Press, 2001.

[155] Hans C. Ohanian, "What is the Principle of Equivalence?", American Journal of Physics 45, pp. 903-909, 1977.

[156] Hans C. Ohanian, "The Energy-Momentum Tensor in General Relativity and in Alternative Theories of Gravitation, and the Gravitational vs. Inertial Mass", arXiv:1010.5557v2 [gr-qc], 2013.

[157] Hans C. Ohanian and Remo Ruffini, *Gravitation and Spacetime*, New York: W. W. Norton & Co., 1994.

[158] Daniele Oriti (ed.), "Approaches to Quantum Gravity: Toward a New Understanding of Space, Time, and Matter", Cambridge: Cambridge University Press, 2009.

[159] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, Reading, MA: Perseus Books, 1995.

[160] J. Brian Pitts, "Absolute Objects and Counterexamples: Jones-Geroch Dust, Torretti Constant Curvature, Tetrad-Spinor, and Scalar Density", Studies in History and Philosophy of Modern Physics 37(2), pp. 347-371, 2006.

[161] J. Brian Pitts, "Gauge-Invariant Localization of Infinitely Many Gravitational Energies from All Possible Auxiliary Structures", General Relativity and Gravitation 42, pp. 601-622, 2010.

[162] J. Brian Pitts, "Space-time Constructivism *vs*. Modal Provincialism: Or, How Special Relativistic Theories Needn't Show Minkowski Chronogeometry", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[163] Henri Poincaré, *Science and Hypothesis*, London and Newcastle-on-Tyne: The Walter Scott Publishing Co. Ltd., 1905.

[164] Eric Poisson and Clifford M. Will, *Gravity: Newtonian, Post-Newtonian, Relativistic*, Cambridge: Cambridge University Press, 2014.

[165] Joseph Polchinski, *String Theory*, Vol. 1, Cambridge: Cambridge University Press, 1998.

[166] Joseph Polchinski, "Dualities of Fields and Strings", Studies in History and Philosophy of Modern Physics 59, pp. 6-20, 2017.

[167] Oliver Pooley, *The Reality of Spacetime*, D.Phil. Thesis, University of Oxford, 2002.

[168] Oliver Pooley, "Points, Particles, and Structural Realism", In D. Rickles, S. French & J. Saatsi (eds.), *The Structural Foundations of Quantum Gravity.* Oxford: Oxford University Press, 2006.

[169] Oliver Pooley, “Substantivalist and Relationist Approaches to Spacetime”, in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, 2013.

[170] Oliver Pooley, *The Reality of Spacetime*, book manuscript, 2015.

[171] Oliver Pooley, “Background Independence, Diffeomorphism Invariance, and the Meaning of Coordinates”, in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, 2017.

[172] R. V. Pound and G. A. Rebka Jr., “Apparent Weight of Photons”, Physical Review Letters 4(7), pp. 337-341, 1960.

[173] Willard Van Orman Quine, “On the Reasons for Indeterminacy of Translation”, Journal of Philosophy 67(6), pp. 178-83, 1970.

[174] Willard Van Orman Quine, “On Empirically Equivalent Systems of the World”, Erkenntnis 9, pp. 313-328, 1975.

[175] James Read, “Background Independence in Classical and Quantum Gravity”, B.Phil. Thesis, University of Oxford, 2016.

[176] James Read, “The Interpretation of String-Theoretic Dualities”, Foundations of Physics 46(2), pp. 209-235, 2016.

[177] James Read, “Explanation, Geometry, and Conspiracy in Relativity Theory”, in C. Beisbart, T. Sauer and C. Wüthrich (eds.), *Thinking About Space and Time: 100 Years of Applying and Interpreting General Relativity*, vol. 15 of the *Einstein Studies* series, Basel: Birkhäuser, 2018. (Forthcoming.)

[178] James Read, “Functional Gravitational Energy”, British Journal for the Philosophy of Science, 2018. (Forthcoming.)

[179] James Read, “On Miracles and Spacetime”, 2018.

[180] James Read, Harvey R. Brown and Dennis Lehmkuhl, “Two Miracles of General Relativity”, Studies in History and Philosophy of Modern Physics, 2018. (Forthcoming.)

[181] James Read and Baptiste Le Bihan, “Duality and Ontology”, Philosophy Compass, 2018. (Forthcoming.)

[182] James Read and Thomas Møller-Nielsen, “Motivating Dualities”, Synthese, 2018. (Forthcoming.)

[183] James Read and Nicholas J. Teh, “The Teleparallel Equivalent of Newton-Cartan Gravity”, Classical and Quantum Gravity 35, 18LT01, 2018.

[184] Dean Rickles, “A Philosopher Looks at String Dualities”, Studies in History and Philosophy of Modern Physics 42, pp. 54-67, 2011.

[185] Dean Rickles, “AdS/CFT Duality and the Emergence of Spacetime”, Studies in History and Philosophy of Modern Physics 44, pp. 312-320, 2013.

[186] Dean Rickles, “Mirror Symmetry and Other Miracles in Superstring Theory”, Foundations of Physics 43, pp. 54-80, 2013.

[187] Dean Rickles, “Dual Theories: ‘Same but Different’ or ’Different but Same’?”, Studies in History and Philosophy of Modern Physics, pp. 62-67, 2017.

[188] Gonzalo Rodriguez-Pereyra, *Leibniz’s Principle of Identity of Indiscernibles*, Oxford: Oxford University Press, 2014.

[189] Carlo Rovelli, “Why Gauge?”, Foundations of Physics 44, pp. 91-104, 2014.

[190] Simon Saunders, “Indiscernibles, General Covariance and Other Symmetries: The Case for Non-Reductive Relationalism”, in A. Ashketar *et al.* (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honour of John Stachel*, Dordrecht: Kluwer Press, 2003.

[191] Simon Saunders, "Physics and Leibniz's Principles", in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, 2003.

[192] Simon Saunders, "Rethinking Newton's *Principia*", Philosophy of Science 80, pp. 22-48, 2013.

[193] Erwin Schrödinger, "Die Energiekomponenten des Gravitationsfeldes", *Physikalische Zeitschrift* 19, pp. 4-7, 1918.

[194] Erwin Schrödinger, *Space-Time Structure*, Cambridge: Cambridge University Press, 1950.

[195] L. I. Schiff, "On Experimental Tests of the General Theory of Relativity, American Journal of Physics 28, pp. 340-343, 1960.

[196] Bernard Schutz, *A First Course in General Relativity*, second edition, Cambridge: Cambridge University Press, 2009.

[197] Mark Srednicki, *Quantum Field Theory*, Cambridge: Cambridge University Press, 2007.

[198] Howard Stein, "Some Philosophical Prehistory of General Relativity", in J. Earman, C. Glymour and J. Stachel (eds.), *Foundations of Space-Time Theories*, Minnesota Studies in the Philosophy of Science, volume 8, Minneapolis: University of Minnesota Press, pp. 3-49, 1977.

[199] Syman Stevens, "The Dynamical Approach as Practical Geometry", Philosophy of Science 82, pp. 1152-1162, 2015.

[200] Syman Stevens, "Regularity Relationalism and the Constructivist Project", British Journal for the Philosophy of Science, 2017. (Forthcoming.)

[201] Andrew Strominger, "Quantum Gravity and String Theory: The Past, the Present, and the Future", talk given at *Strings 2014* in Princeton, NJ, 2014.

[202] Patrick Suppes, "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences", Synthese 12, pp. 287-301, 1960.

[203] John Synge, *Relativity: The General Theory*, Amsterdam: North-Holland Publishing Company, 1960.

[204] Nicholas J. Teh, "Holography and Emergence", in Studies in History and Philosophy of Modern Physics 44, pp. 300-311, 2013.

[205] Nicholas J. Teh, "Galileo's Gauge: Understanding the Empirical Significance of Gauge Symmetry", Philosophy of Science 83(1), pp. 93-118, 2016.

[206] Nicholas J. Teh, "Recovering Recovery: On the Relationship Between Gauge Symmetry and Trautman Recovery", Philosophy of Science 85, pp. 201-224, 2018.

[207] Kip S. Thorne, David L. Lee, and Alan P. Lightman, "Foundations For a Theory of Gravitation Theories", Physical Review D 7, pp. 3563-3578, 1973.

[208] Andrzej Trautman, "Conservation Laws in General Relativity", in L. Witten (ed.), *Gravitation: An Introduction to Current Research*, John Wiley and Sons, 1962.

[209] Andrzej Trautman, "Foundations and Current Problems of General Relativity", in S. Deser & K. W. Ford (eds.), *Lectures on General Relativity*, New Jersey: Englewood Cliffs, Prentice-Hall, 1965.

[210] Bas C. van Fraassen, *The Scientific Image*, Oxford: Oxford University Press, 1980.

[211] Bas C. van Fraassen, *Laws and Symmetry*, Oxford: Oxford University Press, 1989.

[212] Bas C. van Fraassen, *The Empirical Stance*, New Haven, CT: Yale University Press, 2002.

[213] Cumrun Vafa, "Geometric Physics", in *Proceedings of the International Congress of Mathematics, Vol. 1*, G. Fischer and U. Rehmann (eds.), 1998.

[214] Robert Wald, *General Relativity*, Chicago: University of Chicago Press, 1984.

[215] David Wallace, *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*, Oxford: Oxford University Press, 2012.

[216] David Wallace, "Fields As Bodies: A Unified Presentation of Spacetime and Internal Gauge Symmetry", 2015. Available at: arXiv:1502.06539.

[217] David Wallace, "Who's Afraid of Coordinate Systems? An Essay on Representation of Spacetime Structure", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[218] David Wallace, "Fundamental and Emergent Geometry in Newtonian Physics", British Journal for the Philosophy of Science, 2017. (Forthcoming.)

[219] James Owen Weatherall, "On (Some) Explanations in Physics", Philosophy of Science 78(3), pp. 421-447, 2011.

[220] James Owen Weatherall, "On the Status of the Geodesic Principle in Newtonian and Relativistic Physics", Studies in the History and Philosophy of Modern Physics 42(4), pp. 276-281, 2011.

[221] James Owen Weatherall, "The Motion of a Body in Newtonian Theories", Journal of Mathematical Physics 52(3), 032502, 2011.

[222] James Owen Weatherall, "A Brief Remark on Energy Conditions and the Geroch-Jang Theorem", Foundations of Physics 42(2), pp. 209-214, 2012.

[223] James Owen Weatherall, "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?", Erkenntnis 81(5), pp. 1073-1091, 2016.

[224] James Owen Weatherall, "Maxwell-Huygens, Newton-Cartan, and Saunders-Knox Spacetimes", Philosophy of Science 83(1), pp. 82-92, 2016.

[225] James Owen Weatherall, "Understanding Gauge", Philosophy of Science 85(5), pp. 1039-1049, 2016.

[226] James Owen Weatherall, "A Brief Comment on Maxwell(/Newton)[-Huygens] Spacetime", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[227] James Owen Weatherall, "Inertial Motion, Explanation, and the Foundations of Classical Spacetime Theories", in D. Lehmkuhl, G. Schiemann, and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Boston, MA: Birkhäuser, pp. 13-42, 2017.

[228] James Owen Weatherall, "Conservation, Inertia, and Spacetime Geometry", Studies in History and Philosophy of Modern Physics, 2017. (Forthcoming.)

[229] James Owen Weatherall, "Categories and the Foundations of Classical Field Theories", in E. Landry (ed.), *Categories for the Working Philosopher*, Oxford: Oxford University Press.

[230] Hermann Weyl, "Zur Infinitesimalgeometrie: Einordnung der Projektiven und der Konformen Auffasung", Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, Göttingen, pp. 99-112, 1921.

[231] Hermann Weyl, *Mathematische Analyse des Raumproblems*, lecture 3, Berlin, 1923.

[232] Hermann Weyl, *Symmetry*, Princeton, NJ: Princeton University Press, 1952.

[233] Edward Witten, "String Theory Dynamics in Various Dimensions", Nuclear Physics B 443, pp. 85-126,1995.

[234] Edward Witten, "Duality, Spacetime and Quantum Mechanics", Physics Today, pp. 28-33, May 1997.

[235] John Worrall, "Structural Realism: The Best of Both Worlds?", Dialectica 43, pp. 99-124, 1989.

[236] Christian Wüthrich, "To Quantize or Not to Quantize: Fact and Folklore in Quantum Gravity", Philosophy of Science 72, pp. 777-788, 2005.

[237] E. Zampino, "A Brief Study on the Transformation of Maxwell Equations in Euclidean Four-Space", Journal of Mathematical Physics 27, pp. 1315-1318, 1986.