Background Independence in Classical and Quantum Gravity

B.Phil. Thesis

James Read

Merton College, University of Oxford, OX1 4JD, UK

james.read@philosophy.ox.ac.uk

Candidate number: 501670

Word count: 29994

Abstract

Background independence is sometimes claimed to be the defining characteristic of general relativity (GR), and a necessary feature of any candidate quantum extension. In this thesis, we appraise numerous definitions of background independence, before assessing whether various spacetime theories—both classical and quantum—manifest this quality. Our aims are three-fold: (i) to clarify the best possible understanding of background independence; (ii) to assess the extent to which, at the level of classical spacetime theories, background independence is characteristic of GR alone; and (iii) to assess whether this quality is in fact manifest in some of our best candidate quantum gravity theories.

Everything, everything, everything, everything

In its right place

In its right place

In its right place

Right place

Contents

Chapter 1

Introduction

The quality of *background independence* is often claimed to be the defining characteristic of general relativity (GR), and a necessary feature of any candidate quantum extension.[1] Nevertheless, there remain multifarious open questions regarding (i) the definition of background independence; (ii) the relation between background independence and other qualities of GR, such as diffeomorphism invariance; and (iii) the status of both classical and quantum spacetime theories with respect to background independence.

The purpose of this thesis is to resolve these questions. In §2, we present the *semantic conception* of scientific theories – which will prove essential in the ensuing. In §3, we consider four definitions of background independence; their relations to other important qualities of spacetime theories (including GR), such as general covariance and diffeomorphism invariance; their applications to theories such as GR; scope; interrelations; and limitations.

These in hand, in §4 we address the background independence of four classical spacetime theories: Newtonian gravitation (NG), Newton-Cartan theory (NCT); teleparallel gravity

[1]Claims of this form can be found in the works of e.g. Rovelli [113, 114] and Smolin [123, 124]; and (from a somewhat different perspective) in those of de Haro *et al.* [25, 27].

(TPG); and Kaluza-Klein theory (KKT). This work facilitates evaluation of these theories with respect to this quality (thereby allowing comparisons with GR to be made); appraisal of the definitions themselves; and the bringing to light of novel features of these theories.[2]

In §5, we discuss ways in which the above four definitions of background independence may be modified to apply to quantum theories. This done, we assess whether both perturbative string theory (§6) and bulk theories in so-called *holographic dualities* (§7), are background independent. There exists a majority verdict that such *is* the case – meaning that background independence cannot be cited as a reason not to pursue the study of these theories.[3]

Throughout this thesis, questions regarding whether background independence should be considered a desideratum on any spacetime theory are set aside. Rather, we are concerned with the *formal* features of this quality; the extent to which such features are exhibited in notable spacetime theories; and the lessons which appraisals of background independence can teach us regarding those theories.

[2]For example, forms of indeterminism in NG (§4.1.3); different senses of gauge invariance in TPG (§4.2.3); and issues regarding the diffeomorphism group associated to KKT (§4.3.2).

[3]See e.g. [124, §5.4].

Chapter 2

Models and Gauge

Throughout this thesis, we embrace the *semantic conception* of scientific theories—introduced by Suppes [126], and famously endorsed by van Fraassen [134, 135]—on which a given theory is associated with a class of *models*.[1] In this chapter, we introduce in §2.1 three important varieties of model; before in §2.2 elaborating on how *gauge redundancies* fit into this framework. Discussion at this stage is focussed exclusively upon classical theories; consideration of the quantum is deferred to §5.

2.1 Three Classes of Model

For a given theory $\mathcal{T}$, we 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.[2]

[1] Van Fraassen identifies a model of a theory as "Any structure which satisfies the axioms of [that] theory" [134, p. 53]. In the language of this thesis, it is natural to identify models in van Fraassen's sense with *dynamically possible models* – or perhaps with the intersection of *dynamically possible models* and *boundary possible models* (see §2.1). In this work, following e.g. [103, 104], we understand the notion of a model in a broader sense, discussed below.

[2] Including tensors, but also (in principle) tensor densities, pseudotensors, etc. Though discussion of such non-tensorial objects shall mostly be set aside in this thesis, for arguments for taking these objects seriously,

For example, the KPMs of general relativity (GR) are picked out by all triples of the form $\langle M, g_{ab}, \Phi \rangle$,[3] where M is a four-dimensional differentiable manifold;[4,5] 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 subspace $\mathcal{D} \subset \mathcal{K}$, the *dynamically possible models* (DPMs) of $\mathcal{T}$. For example, in the case of GR, only those triples $\langle M, g_{ab}, \Phi \rangle$ the geometrical objects of which satisfy the *Einstein field equations*[6]

$$G_{ab} = 8\pi T_{ab} \tag{2.1.1}$$

—the dynamical equations of the theory, which relate g_{ab} to the stress-energy tensor T_{ab} of the Φ[7]—in *addition* to the dynamical equations of the Φ, are DPMs.[8,9]

Though the notions of KPMs and DPMs of a theory are known in the literature (see

see [95, 97].

[3]Throughout this thesis, abstract (i.e. coordinate-independent) indices are written in Latin script; indices in a coordinate basis are written in Greek script; 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. We also set $G_N = c = 1$.

[4]Note with Earman and Norton that one should avoid making *ab initio* "unnecessary global assumptions" regarding manifold topology. See [38, p. 5].

[5]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 (cf. footnote 9). Indeed, in light of the debate over the hole argument [38], it is *not* necessarily correct to interpret M as representing substantial spacetime at all – though this issue shall be set aside in this work.

[6]Barring discussion in footnotes 17, 18 and 20 of §3.3.3, as well as in §7, in this thesis we restrict to the case of vanishing cosmological constant Λ. For $\Lambda \neq 0$, the field equations read $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$.

[7]Sometimes, the matter fields Φ and their associated stress-energy tensor T_{ab} are conflated (see e.g. [55, p. 191], [74, p. 1014]). This is a mistake, and should be avoided – for further discussion, see §3.5.4.

[8]Strictly, independence of these dynamical equations from the Einstein field equations depends on the case in question – for discussion of this point, see [19, §9.3] and [82, §20.6].

[9]Note that we take (following e.g. [15, 105]) the class $\mathcal{K}$ of KPMs of $\mathcal{T}$ to be the class of all possible *histories* of the fields $O_1, \dots, O_n$ (rather than as e.g. instantaneous states). If a choice of such histories is picked out as a solution set of the specified the dynamical equations of the theory, then the associated model is a DPM. On this understanding, one can proceed to interpret models as corresponding to *possible worlds* – see §2.2.

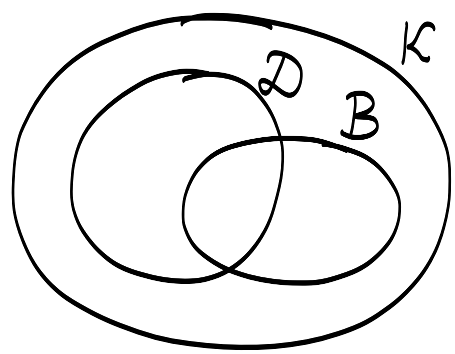

Figure 2.1: The relation between $\mathcal{K}$,$\mathcal{D}$, and $\mathcal{B}$ for a generic $\mathcal{T}$.

e.g. [103,104]), we introduce in this thesis a third class of models of a theory $\mathcal{T}$, which we call the *boundary possible models* (BPMs) $\mathcal{B} \subset \mathcal{K}$. These are those KPMs of $\mathcal{T}$ the O_i of which satisfy certain specified boundary conditions. Again taking the example of GR, sometimes we restrict attention to KPMs of this theory for which the metric field g_{ab} satisfies the boundary condition of *asymptotic flatness* (see e.g. [138, ch. 11], and §§3.4.2, 3.5.3, and 7.1).

Where a class $\mathcal{B}$ is specified for $\mathcal{T}$, we may take those models to specify the *KPMs* of a restricted theory $\mathcal{T}'$. In that case, $\mathcal{T}'$ is a *sub-theory* of the *super-theory* $\mathcal{T}$.[10] Finally, for a given $\mathcal{T}$ with specified dynamical equations and boundary conditions on its geometrical objects, there is in general no restriction on the overlap between $\mathcal{D}$ and $\mathcal{B}$ – the generic picture thus being as per Fig. 2.1.

2.2 Gauge Redundancies

Models of a theory $\mathcal{T}$ may be *interpreted* as representing possible worlds.[11] 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 rep-

[10] The significance of this move will become clear in §§3.4.2 and 7.3.1.

[11] See e.g. [105, §2].

resenting the same world is an example of a *gauge redundancy*. In the case in which the interpretation of $\mathcal{T}$ leads to gauge redundancy, we may construct a *reduced* space of models $\tilde{\mathcal{K}}$, in which gauge-related models are *identified*.[12]

The above is purely formal; there remains an outstanding question concerning *when* two models of $\mathcal{T}$ should be interpreted as representing the same world. One popular line (found in e.g. [31, 50, 119, 120]) is as follows. First, define a *symmetry transformation* to be a transformation upon the O_i in the KPMs of $\mathcal{T}$, such that DPMs of $\mathcal{T}$ are always taken to DPMs.[13,14] The claim is, then, that two models of $\mathcal{T}$ should be regarded as representing the same possible world if they are related by a symmetry transformation.[15]

According to this line on symmetries, which Møller-Nielsen dubs the *interpretational* approach [83, p. 4], in the presence of symmetry-related models, we may first (a) interpret those models as representing the same possible world; then (b) *identify* such models, to construct a reduced space of KPMs $\tilde{\mathcal{K}}$; and *finally* (c) and seek to provide a "metaphysically perspicuous characterisation" of the models of $\tilde{\mathcal{K}}$.[16] This is in contrast with the *motivational* approach [83, p. 4], according to which the existence of symmetry-related models first (a) *motivates* us to

[12]For a concise expression of these points in the language of category theory, see [141–143].

[13]For an example of such a symmetry transformation, consider e.g. uniform boosts in Newtonian gravitation (NG) – this theory being discussed in more detail in §4.1.2.

[14]In fact, as Belot points out, such a definition of a symmetry transformation is too broad: "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 size of a theory's group of symmetries depends only on the size of its space of solutions" [15, p. 6]. Though Belot considers more nuanced definitions of symmetry transformations [15, §§ 3-4] (which, for simplicity, are set aside in this work), he argues that all such definitions come into conflict with what Møller-Nielsen calls the 'interpretational approach to symmetries' (discussed below) [83, p. 4], according to which symmetry-related models represent the same possible world. This lends further weight to Belot's claims that the interpretational account should be rejected – see below.

[15]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. For a discussion of such issues, see [105, §2]; in this paper, we 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.

[16]For discussion of what such a "metaphysically perspicuous characterisation" consists in, see [83, §3]. Note that some advocates of the interpretational approach deny that the seeking of such a characterisation is necessary at all – see e.g. [31, p. 322].

provide a "metaphysically perspicuous characterisation" of the shared ontology of these models; but only once such a characterisation is constructed should we (b) interpret those models as representing the same possible world; and (c) identity those models to construct a reduced space of KPMs, $\tilde{\mathcal{K}}$.

In line with Møller-Nielsen [83, p. 3], Belot argues that in several important circumstances, symmetry-related models represent *distinct* physical possibilities [16, p. 4]. One such case is that of models of GR upon which asymptotic boundary conditions are imposed:[17,18] [13, p. 3] (Cf. also [16, pp. 29ff.].)

> Consider a sector of GR in which asymptotic boundary conditions have been imposed. Let the space of DPMs be denoted by $\mathcal{D}$, the group of spacetime diffeomorphisms that preserve the asymptotic boundary conditions be denoted by D, and the group of spacetime diffeomorphisms asymptotic to the identity at infinity by D_0. In typical cases of physical interest, we find that D can be viewed as the product of D_0 with a group G that is called the *asymptotic symmetry group* because it can be thought of as acting geometrically at infinity. One then expects to find that G acts also on $\mathcal{D}/D_0$ and that this action is associated in the usual way with the conserved quantities that the imposition of asymptotic boundary conditions bring into existence. In this case, it is natural to take solutions to be gauge equivalent if and only if related by a diffeomorphism in D_0 so that the space $\mathcal{D}/D_0$ can be viewed as representing the physics of the theory without redundancy. But [the interpretational approach] would appear to recommend instead taking the space $\mathcal{D}/D$ to play this role—which would efface the interesting representation of G as the symmetry group of the theory.

[17] In this quotation, notation has been altered for consistency with this thesis; there is no change in content.

[18] Whether Belot should be classified as endorsing the motivational approach on the basis of this example is, however, unclear, for it appears that in this case Belot does not even consider us *motivated* to find a "metaphysical perspicuous characterisation" of the shared ontology of symmetry-related models.

In light of such cases,[19] there exist legitimate reasons to question the interpretational approach, and to take seriously the motivational alternative. Wherever one stands in this debate, however, the important point to note is that Belot does not subscribe to the former programme. Indeed, counting the number of physically inequivalent models which are related by certain symmetries is a crucial aspect of his account of background independence (see §3.5).

[19] And also—perhaps ultimately more important—the issues raised in footnotes 14 and 15.

Chapter 3

Classical Background Independence

In this chapter, we discuss the concepts of *general covariance* (§3.1), *diffeomorphism invariance* (§3.2), and *absolute objects* (§3.3). These introduced, in §3.3.3 we assess the possibility of defining *background independence* as the absence of absolute objects in a theory, before in §3.4 considering two definitions of background independence due to Pooley [104]. In §3.5, we discuss a fourth definition, due to Belot [14, p. 2877]. For all of these approaches to background independence, we highlight potential problems and suggest associated revisions; we close with an integrated discussion in §3.6.

3.1 General Covariance

The first concept in need of discussion is *general covariance*. We follow Pooley in giving the following definition: [104, p. 12]

Definition 1. (General covariance) *A formulation of a theory is* generally covariant *iff the equations expressing its laws are written in a form that holds with respect to all members of a* *set of coordinate systems that are related by smooth but otherwise arbitrary transformations.*

With such a definition in mind, Friedman states that "the principle of general covariance has no physical content whatever: it specifies no particular physical theory; rather it merely expresses our commitment to a certain style of formulating physical theories" [46, p. 55]. The reason for this—originally pointed out by Kretchmann in a communication with Einstein in 1917 [69]—is that *any* theory may be given a generally covariant formulation.[1] Thus, GR does not differ from special relativity (SR)—or any other spacetime theory—in virtue of having a generally covariant formulation.

As Pooley points out, however, GR *does* differ from SR in lacking a *non*-covariant formulation [104, p. 9]. This lack of preferred coordinates is due to the fact that the spacetime structures of a generic solution of the theory lack symmetries and so cannot be encoded in special coordinates. This lack of symmetry is entailed by, but does not entail, dynamical spacetime structure. A fully dynamical field, free to vary from solution to solution, will in general lack symmetries. The converse is not true, for in principle we can define a theory involving a fixed metric with no isometries, and such a theory will only have a generally covariant formulation [104, p. 9].[2]

3.2 Diffeomorphism Invariance

3.2.1 Diffeomorphisms

Let us now consider the notion of *diffeomorphism invariance*. For a given spacetime theory $\mathcal{T}$ with arbitrary KPM $\mathcal{M} = \langle M, O_1, \ldots, O_n \rangle$, consider a mapping d of the O_i on M, taking $\mathcal{M}$

[1] For further discussion, see [5, 46, 85, 140].

[2] Smolin calls into question this claim [123, p. 201]. We endorse Pooley's response, found at [104, fn. 17].

to a new model $d^*\mathcal{M} := \langle M, d^*O_1, \ldots, d^*O_n \rangle$. If d is smooth, one-one, onto, and has smooth inverse, then it is called a *diffeomorphism*. These diffeomorphisms form a group, denoted Diff (M).[3]

3.2.2 Two Theories

To discuss *diffeomorphism invariance*, it is useful to introduce two exemplar spacetime theories. The first—a special relativistic theory with KPMs $\langle M, \eta_{ab}, \varphi \rangle$, which, following Pooley [104, p. 13], we refer to as **SR1**—is a theory of a real scalar field φ, with DPMs picked out by the massless Klein-Gordon equation[4]

$$\eta^{ab}\varphi_{;ab} = 0. \tag{3.2.1}$$

In the KPMs of **SR1**, the Minkowski metric field η_{ab} is *fixed* – i.e. is identically the same in every model. The dynamically possible models (DPMs) are then the proper subset of the KPMs picked out by the requirement that φ satisfies (3.2.1) relative to the η_{ab} common to all the KPMs. So understood, (3.2.1) is not an equation for η_{ab} and φ together. Rather, it is an equation for φ alone, *given* η_{ab} [104, p. 13].

Now consider an exemplar general relativistic theory – call it **GR1** [104, p. 14]. Here, KPMs are triples $\langle M, g_{ab}, \varphi \rangle$; in this case, the Lorentzian metric field g_{ab} is no longer fixed, and the DPMs are picked out as a subset of the KPMs by two equations – the field equations

[3] For details, see e.g. [138, pp. 437ff.].

[4] In **SR1**, a (unique) torsion-free, metric-compatible derivative operator is to be defined in terms of the metric field η_{ab}, and is not another primitive object.

(2.1.1), and the analogue of (3.2.1),

$$g^{ab}\varphi_{;ab} = 0. \tag{3.2.2}$$

Here, (3.2.2)—unlike (3.2.1)—is not an equation for φ *given* g_{ab}. Rather, (3.2.2) and the EFEs form a coupled system of equations for g_{ab} and φ *together* [104, p. 14].[5]

Fixed fields such as the η_{ab} of **SR1** are standardly labelled *background fields*. In the foundational literature on relativity, it is sometimes claimed that the essential novelty of GR is that such background fields have been excised from the theory – so that GR qualifies as *background independent* [104, pp. 2-3]. In §3.2.4, we assess whether this is correct.

3.2.3 Diffeomorphism Invariance

With the above in hand, we are in a position to discuss diffeomorphism invariance.[6] As Pooley notes [104, p. 13], one common such definition—especially in the post-hole argument foundational literature (see e.g. [34, p. 47])—is the following:

Definition 2. (Diffeomorphism invariance) *A theory $\mathcal{T}$ is* diffeomorphism invariant *iff, if $\langle M, O_1, \ldots, O_n \rangle$ is a DPM of $\mathcal{T}$, then so is $\langle M, d^*O_1, \ldots, d^*O_n \rangle$, for all $d \in \mathrm{Diff}\,(M)$.*[7]

When applied to **SR1**, Def. 2 yields the verdict that the theory is *not* diffeomorphism

[5]Discussion of the distinction between theories with fixed fields—such as **SR1**—and theories with no such fixed fields—such as **GR1**—can also be found in the works of Giulini [49, p. 114] and Vassallo [137, pp. 5ff.]. These works also discuss theories akin to Pooley's **SR2** [104, p. 18], introduced in §3.2.4.

[6]While in this subsection we focus exclusively on diffeomorphism invariance, the setup could in principle be extended to any other symmetries of a certain theory.

[7]In fact, this is often taken as the definition of general covariance (see e.g. [46, pp. 51ff.]). As Pooley states, "In arguing for this equivalence, [such authors appear] to overlook the crucial possibility ... that a coordinate-free equation relating two geometric objects A and B, can nonetheless be interpreted as an equation for B alone, given a fixed A" [104, p. 13]. This drives a wedge between the notions of general covariance and diffeomorphism invariance, and is discussed further both in this subsection and in §3.2.4 below.

invariant. The reason is that Def. 2 requires that if $\langle M, \eta_{ab}, \varphi \rangle$ is a model of the theory, then so is $\langle M, d^*\eta_{ab}, d^*\varphi \rangle$, for an arbitrary diffeomorphism d. However, this definition can only be satisfied when $d^*\eta_{ab} = \eta_{ab}$; for the diffeomorphisms which do *not* satisfy this condition, the model $\langle M, d^*\eta_{ab}, d^*\varphi \rangle$ is not even a KPM of the theory, let alone a DPM [104, p. 14]. What goes for **SR1** here also applies *mutatis mutandis* for any other theory with fixed fields – so, on Def. 2, no such theory can be diffeomorphism invariant.

One might question at this stage, however, whether one should consider unrestricted diffeomorphisms in Def. 2, for a *prima facie* distinct and interesting question is the following: *given* that a diffeomorphism keeps us within the KPMs of the theory, does it also keep us within the DPMs? Following this line, one may choose to restrict in Def. 2 to diffeomorphisms which keep us within the space of KPMs of the theory – and then ask whether such transformations also keep us within the space of DPMs. For example, in **SR1**, on this approach we automatically restrict to diffeomorphisms satisfying $d^*\eta_{ab} = \eta_{ab}$, and then ask whether the transformed model $\langle M, d^*\eta_{ab}, d^*\varphi \rangle = \langle M, \eta_{ab}, d^*\varphi \rangle$ is also a DPM of the theory.

Such an approach, however, does not deliver the correct verdict on the diffeomorphism invariance of theories such as **SR1**. For example, in this theory, the condition $d^*\eta_{ab} = \eta_{ab}$ restricts to diffeomorphisms which act as Lorentz transformations on the geometric objects of the theory. But since the dynamical equation of **SR1** (3.2.1) is Lorentz invariant, it is clear that if $\langle M, \eta_{ab}, \varphi \rangle$ is a DPM of the theory, then so too will be $\langle M, d^*\eta_{ab}, d^*\varphi \rangle$ for d so restricted – delivering the intuitively incorrect verdict that **SR1** *is* diffeomorphism invariant.

Thus, we see that restricting the scope of d in Def. 2 *à la* the above does not deliver a plausible definition of diffeomorphism invariance. One way out here—which captures our desire to assess whether transformations which keep us within the space of KPMs of a theory also keep us within its space of DPMs—is due to Pooley [104, p. 13]. Letting F_i stand for a fixed field common to all KPMs (such as η_{ab} in **SR1**), and D_i stand for a dynamical field, Pooley defines diffeomorphism invariance as follows: [104, p. 15]

Definition 3. (Diffeomorphism invariance, final) *A theory $\mathcal{T}$ is* diffeomorphism invariant *iff, if $\langle M, F_1, \ldots, F_n, D_1, \ldots, D_m \rangle$ is a DPM of $\mathcal{T}$, then so is $\langle M, F_1, \ldots, F_n, d^* D_1, \ldots, d^* D_m \rangle$, for all $d \in \text{Diff}(M)$.*[8]

Note that on Def. 3, **SR1** fails to be diffeomorphism invariant, as required: for an arbitrary diffeomorphism, if $\langle M, \eta_{ab}, \Phi \rangle$ is a solution of **SR1**, then $\langle M, \eta_{ab}, d^* \Phi \rangle$ in general will not be. However, the diffeomorphism invariance of theories with fixed fields is no longer precluded by fiat – rather, it will fail in virtue of the transformed fields no longer satisfying the dynamical equations of the theory in question (thus also avoiding the pitfalls of our above proposed modification of Def. 2). This is a clear advantage of Def. 3. In addition, note that **GR1** is diffeomorphism invariant on unrestricted Def. 2: if $\langle M, g_{ab}, \varphi \rangle$ is a model of the theory, then so is $\langle M, d^* g_{ab}, d^* \varphi \rangle$, for arbitrary d. Since this is the case, **GR1** is also diffeomorphism invariant on restricted Def. 2 (in which we consider only diffeomorphisms which keep us within the KPMs of the theory), and Def. 3.

In the remainder of this thesis, we fix the sense of 'diffeomorphism invariance' to be that of Def. 3. One important upshot is that general covariance is not identical with diffeomorphism invariance. To illustrate, the dynamical equations of **SR1** are generally covariant; nevertheless, the theory is *not* diffeomorphism invariant. Several authors err in making such an identification – for example, Samaroo writes that "the requirement of general covariance is satisfied to the extent that all of a theory's geometric objects are invariant under $\text{Diff}(M)$; thus, I will refer henceforth to general covariance as diffeomorphism-invariance" [118, p. 1075].[9]

[8] This is a special case of the definition of a dynamical symmetry found at [34, p. 45]. Cf. footnote 6.

[9] In fact, this statement constitutes a more serious error, for invariance of a theory's geometric objects under diffeomorphisms is surely too strong a criterion for diffeomorphism invariance – rather, the relevant question is whether the transformed fields still satisfy the theory's dynamical equations.

3.2.4 Diffeomorphism Invariance and Background Independence

Does diffeomorphism invariance coincide with background independence? Intuitively no, for consider a theory of a real scalar field φ with models $\langle M, g_{ab}, \varphi \rangle$ and dynamical equations

$$g^{ab}\varphi_{;ab} = 0, \tag{3.2.3}$$

$$R^{a}{}_{bcd} = 0. \tag{3.2.4}$$

Here, $R^{a}{}_{bcd}$ is the Riemann curvature tensor associated to the metric field g_{ab}.[10] Once more following Pooley [104, p. 18], call this theory **SR2**.[11,12] **SR2** is diffeomorphism invariant on Def. 3. Nevertheless, it is intuitively *not* background independent, for although the metric field is not fixed in all KPMs as with **SR1**, it *is* fixed[13] in all DPMs via (3.2.4). This indicates that diffeomorphism invariance cannot be equated with background independence.

3.3 Absolute Objects

3.3.1 Andersonian Absolute Objects

Anderson, echoing Kretschmann's observation that any theory can be given a generally covariant formulation, challenges the view that general covariance is the characteristic feature of

[10] Via the associated (unique) torsion-free, metric-compatible derivative operator – cf. footnote 5.

[11] Note that unlike **SR1**, the metric field of **SR2** is *not* fixed *ab initio*.

[12] One might complain that neither **SR1** nor **SR2** is the most natural formulation of a special relativistic theory of the Klein-Gordon field – rather, KPMs of such a theory should be those constructed by *quotienting* the $\mathcal{K}$ of **SR2** by models the metric fields of which are related by diffeomorphisms. Since, as for Pooley, in this thesis **SR1** and **SR2** are *toy models* introduced for illustrative purposes, this issue shall here be set aside.

[13] Or, to be more precise, is equivalent up to isomorphism.

GR [4, 5]. He claims, rather, that this distinguishing feature consists in GR's representation of its metric affine geometry by a *dynamical object* [5, p. 83], where such dynamical objects are defined by way of contrast with so-called *absolute objects*. Anderson characterises absolute objects as follows: *objects the same in all DPMs*;[14] for our purposes, it suffices to elevate this to a definition:

Definition 4. (Absolute object) *A geometrical object the same (up to isomorphism) in all DPMs of a theory.*

The complement of the class of absolute objects in the class of the geometric objects in the models of a theory is the class of dynamical objects [5, p. 83]. Sometimes, it is claimed (see e.g. [95]) that a theory is *background independent* iff it has no absolute objects in its formulation. We assess this proposal in §3.3.3; here we again elevate it to a definition:

Definition 5. (Background independence, absolute objects) *A theory is* background independent *iff it has no absolute objects in its formulation.*

3.3.2 Confined Objects

It is worth mentioning one other class of objects which may appear in the models of a theory – referred to by Pitts as *confined objects* [95, §2] (cf. [130]). Confined objects are structures (not necessarily geometrical objects, in the sense of §2.1) which do not change *at all*, under any coordinate transformation [95, p. 354]. As Pitts states, examples include the identity matrix, the Lorentz matrix $\mathrm{diag}\,(-1, 1, 1, 1)$, fixed Dirac γ^{μ} matrices, Lie group structure constants, and the Kronecker delta and Levi-Civita alternating symbol. Other possible examples of confined

[14] For a more formal presentation of Anderson's account, also commenting on the differences between Anderson's original presentation [5, p. 83] and Friedman's later reformulation [46, pp. 58-60], see [95, pp. 348-353].

objects include the metric signature [18];[15] and the global topology of spacetime [118, 123].[16] Smolin implies in [123, p. 204] that all such confined objects should ultimately be excised from our final theories of physics – exploring the extent to which this is both true and possible is an interesting task for future pursuit.

3.3.3 Background Independence and Absolute Objects

As mentioned in §3.3.1, sometimes background independence is equated with the absence of absolute objects in a theory. However, as noted by Pitts [95], Giulini [49], Pooley [104], and Sus [127, 128], there are problems with this suggestion: [104, pp. 23-24]

1. There appear to exist cases in which structure that intuitively should count as background is not classified as absolute.
2. There appear to exist cases in which structure that intuitively should not count as background is classified as absolute.
3. On Anderson's definition, GR itself apparently turns out to have an absolute object (and so should count as background dependent).

Let us review each of (1)-(3) in turn.

Case 1: Torretti [131] gives an example falling into this class, considering a theory of modified Newtonian kinematics in which each model's spatial metric has constant curvature, but different models have different values of that curvature. Because the spatial metric in every model

[15] Brown notes that the metric signature in GR is not a consequence of the field equations [18]. Indeed, in principle (as e.g. Gibbons and Hartle have discussed [48]), metric signature *change* is possible in GR (or some closely-related theory). If this work is vindicated, then metric signature may cease to be a confined object.

[16] Mirror symmetry in string theory seems to offer a case where spacetime topology ceases to be fundamental – for a philosophically-oriented introduction, see [110].

has constant curvature, "such a theory surely has something rather like an absolute object in it" [95, p. 363]. Nevertheless, the failure of the metrics to be locally diffeomorphically equivalent for distinct curvature values entails that the metric tensor does not satisfy Anderson's definition of an absolute object.

What should be made of this example? Arguably, it does not succeed, for, as Pitts observes [95, pp. 17-18], if one decomposes the spatial metric into a conformal spatial metric density and a scalar density, then the former *is* an absolute object (cf. (3) below), while the latter, while constant in space and time, counts as a genuine, global degree of freedom [95, p. 18]. Thus, this particular 'counterexample' to the equation of background independence with the absence of absolute objects does not go through. (This, of course, does not preclude the formulation of other counterexamples falling into this first case.)

Case 2: The best-known example of this type is that of Jones-Geroch 'dust'. Consider a general relativistic theory in which the metric field is coupled to matter characterised only by a 4-velocity field U^{μ} and a mass density. Then, as Friedman states, "since any two timelike, nowhere-vanishing vector fields defined on a relativistic space-time are d-equivalent, it follows that any such vector field counts as an absolute object ... and this is surely counter-intuitive" [46, p. 59]. In response to this example, Pitts, following Friedman, argues that such a 'dust' field violates a ban laid down by Anderson on physically redundant variables [95, pp. 355-356, 361-362], writing that "the proscription of irrelevant variables serves to eliminate the Jones-Geroch counterexample: the dust 4-velocity U^{μ} does not count as an absolute object for GTR + dust because U^{μ} does not exist where there is no dust" [95, p. 356].

As Pooley states, however, it is unclear whether this response gets to the heart of this counterexample, for "[i]n the context of this theory, the non-vanishing velocity field is, intuitively, as dynamical as the the 4-momentum. The trouble arises not because we mistook as indispensable an object that Anderson's definition correctly classifies as absolute. The trouble is that Anderson's definition, intuitively, misclassifies that object" [104, p. 24]. In light of this, Pooley suggests that the notion of absolute objects might not, in fact, be a better candidate than the notion of fixed fields for articulating the sense of 'dynamical' relevant to characterising background structure, for no such nowhere-vanishing velocity field qualifies as a fixed field, even if it seemingly *may* count as an Andersonian absolute object.

Case 3: A similar strategy might be pursued in the third case, which regards Pitts' claim that GR itself has an absolute object [95, §8]. To see how Pitts reasons to this conclusion, first recall that in so-called *unimodular general relativity* (UGR),[17] since a metric tensor as a geometrical object is decomposable into a conformal metric density and a scalar density, the unimodular coordinate condition $\sqrt{-g} = 1$ means that there exists an absolute object in this theory.[18] Once such an example is noticed in the context of UGR, Pitts points out that, in fact, similar reasoning holds for GR itself: [95, p. 366]

> GR has an absolute object! This absolute object [$\sqrt{-g}$] is a scalar density of nonzero weight, because every neighbourhood in every model spacetime admits coordinates (at least locally) in which the component of the scalar density has a value of -1.

As Pooley states [104, p. 25], one might accept this verdict without accepting that this automatically means that GR should count as background dependent.[19] Such a move is possible if, for example, one rejects the equation of background independence and the absence of absolute objects, and, rather, attempts to return to the view according to which the absence of *fixed fields* (in the sense of **SR1**) is somehow connected with background independence. In that

[17] That is, the version of GR with cosmological constant—for which the field equations read $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$, and the associated Lagrangian is $\mathcal{L} = \mathcal{L}_G + \mathcal{L}_M$, with $\mathcal{L}_G = \int_M d^4x\sqrt{-g}\,(R - 2\Lambda)$ and $\mathcal{L}_M$ the matter Lagrangian—in which the so-called *trace-free Einstein equations*, $R_{ab} - \frac{1}{4}Rg_{ab} = T_{ab} - \frac{1}{4}Tg_{ab}$, are derived as the extrema of $\mathcal{L}$ via limitation to variations which satisfy the *unimodular coordinate condition* $\sqrt{-g} = 1$ [35, p. 568].

[18] As Pitts mentions [95, p. 366], there also exists a second version of UGR which admits any coordinates, with the help of a non-variational scalar density; this version of UGR clearly also possesses an absolute object.

[19] For an alternative approach, see [127, ch. 3], in which Sus argues for a non-standard definition of *invariance* under a transformation. On this definition, it turns out that GR does *not* have an absolute object.

case, for the background independence of GR, one would require that $\sqrt{-g}$ be interpretable as a fixed field – which is not correct. Pooley explores this option in two further definitions of background independence [104], to which we turn in §3.4.[20]

The upshot of the above three cases is that while some 'counterexamples' to the equation of the background independence of a spacetime theory with its lack of absolute objects do not go through, to the extent that these *are* successful, the problems they present can be overcome by equating background independence not with the absence of absolute objects, but with the absence of *fixed fields*. As we have seen in §3.2.4, however, *prima facie* there exist problems with this proposal: for example, it cannot account for the background dependence of **SR2**. In the following subsection, we see how such problems may be overcome.

3.4 Fixed Fields and Variational Principles

There exist problems with the equation of background independence and the absence of absolute objects. Accordingly, in this subsection we turn to two definitions of background independence due to Pooley [104], both of which constitute reappraisals of the link between background independence and the absence of *fixed fields* (in the sense of **SR1**).

[20] In connection with UGR, it is worth remarking on some comments made by Earman [35]. In GR with non-zero cosmological constant, Λ is non-dynamical. With this in mind, Earman states that "if Λ is not a universal constant, it too must be varied; and the consequence of this variation of [$\mathcal{L} = \mathcal{L}_G + \mathcal{L}_M$ with $\mathcal{L}_G = \int_M d^4x\sqrt{-g}\,(R - 2\Lambda)$ and $\mathcal{L}_M$ the matter Lagrangian] is the absurd result that $\sqrt{-g} = 0$, i.e. that the volume of spacetime is zero" [35, p. 562]; from this, he infers that Λ must be a universal constant, i.e. constant across all models (KPMs). By contrast, Earman continues, in the UGR formalism, the cosmological constant arises as an integration constant from $\nabla_a\,(R + T) = 0$, which follows from the UGR field equations and the energy-momentum conservation equation $\nabla^a T_{ab} = 0$, the latter of which in this formalism (and unlike GR) does *not* follow from the field equations, but rather must be postulated separately [35, p. 563]. For this reason, in UGR Earman claims that Λ *is* free to vary across models (sometimes, this is claimed to suggest a solution to the infamous *cosmological constant problem* – for sceptical discussion, see [35, §4]). What we wish to point out here is that it is not clear why a non-variational (i.e. fixed) field such as Λ in GR may not differ from model to model (and yet, nevertheless, be *fixed* in each model) – in fact, we see in §3.5.4 that this possibility poses problems for Belot's definition of background independence.

3.4.1 Fixed Fields

At [104, pp. 20, 23], Pooley hints at the following definition of background independence:

Definition 6. (Background independence, fixed fields) *A theory is* background independent *iff it has no formulation which features fixed fields.*

The idea here is that while, for example, **SR2** does not feature any fixed fields, it has an alternative formulation—**SR1**—which *does* feature fixed fields; for this reason, **SR2** fails to be background independent. Def. 6 appears to deliver intuitively correct results. To be satisfactory, however, some criterion must be provided for when two *prima facie* distinct theories are, in fact, mere reformulations of the same theory. Drawing upon [104, p. 20], one might specify this criterion to consist in the following:

Definition 7. (Equivalent theory formulations) *Two theories, $\mathcal{T}_1$ and $\mathcal{T}_2$, are* equivalent formulations *of the same theory when:*

1. *KPMs of $\mathcal{T}_1$ and $\mathcal{T}_2$ involve the same types of geometric object.*
2. *DPMs of $\mathcal{T}_1$ and $\mathcal{T}_2$ are isomorphic, up to classes of diffeomorphism-related models.*

In Def. 7, (1) can be cashed out as follows: if KPMs of $\mathcal{T}_1$ consist in a certain number of (r, s)-tensors, weight w tensor densities, etc., then so too must the KPMs of $\mathcal{T}_2$. Such a criterion is clearly satisfied in the case of **SR1** and **SR2**, in spite of the metric field η_{ab} being a fixed field in the former case.[21] On (2), note that in the case of **SR1** and **SR2**, DPMs of the former theory are isomorphic to DPMs of the latter; it is just that, for each DPM of **SR1**, there is an infinite set of DPMs of **SR2** which differ by a diffeomorphism on the metric field. Nevertheless, as Pooley states, "in the sense that matters, the metric structure of spacetime

[21] Formulating criterion (1) in this manner excludes many theories related by *dualities* (see §7) from qualifying inappropriately as theory reformulations.

[in **SR2**] does not differ from DPM to DPM: the g_{ab} in any two DPMs are isomorphic to one another" [104, p. 20]. It is *this fact* which motivates our more lenient formulation of (2) than mere isomorphism between the respective $\mathcal{D}$ of $\mathcal{T}_1$ and $\mathcal{T}_2$.

Clearly,[22] with Def. 7 in hand, Def. 6 delivers the correct verdict that both **SR1** and **SR2** are background dependent. One problem with such a definition, however, is that one cannot straightforwardly use it to assess the background independence of a theory *on its own terms* – and absent any argument that there exist *no* formulations of that theory which feature fixed fields, it can be difficult to appraise using Def. 6 whether or not the theory is indeed background independent. Nevertheless, Def. 6 remains valuable; we see the substantive work to which it can be put in the following chapters.

3.4.2 Variational Principles

In spite of **SR2** and **GR1** sharing the same space of KPMs, there exists an important formal difference between the two theories [104, p. 27]. In the case of **GR1**, the space of DPMs picked out by its dynamical equations can be derived from a variational principle via the action

$$S_{\mathrm{GR1}} = \int d^4x \left(\mathcal{L}_G + \mathcal{L}_\varphi \right), \tag{3.4.1}$$

where

$$\mathcal{L}_G = \sqrt{-g}\kappa R \tag{3.4.2}$$

[22] And, indeed, almost by construction – albeit on independently plausible grounds.

is the Einstein-Hilbert 'gravitational' part of the Lagrangian (R being the Ricci scalar and κ a suitable constant), and

$$\mathcal{L}_\varphi = \sqrt{-g} g^{ab} \varphi_{;a} \varphi_{;b} \tag{3.4.3}$$

is the 'matter' Lagrangian of the massless Klein-Gordon field. By contrast, the same is *prima facie* not true of **SR2**. This motivates the idea that background independence might be synonymous with all of a theory's dependent variables being subject to Hamilton's principle [104, p. 28]. In turn, this rules **SR1** from being background independent, since in this case only φ, and not η_{ab}, is subject to Hamilton's principle (in general, this will preclude any theory with fixed fields from being background independent, in line with the morals of §3.3.3). It will also rule out **SR2**—as desired—if the solution space of this theory really is not obtainable from an appropriately formulated action principle.

As noted by both Pitts [95, p. 357] and Pooley [104, p. 28], however, it *is* possible to derive the dynamical equations of **SR2**, (3.2.3) and (3.2.4), from an action principle. In [111], Rosen shows that these can be derived from an action in which the $\mathcal{L}_G$ piece in S_{GR1} is replaced with a different term, $\mathcal{L}_S = \sqrt{-g}\Theta^{abcd} R_{abcd}$. This prescription involves a Lagrange multiplier field Θ^{abcd}, in addition to the fields common to **SR2** and **GR1**. In this new action, all the dependent variables are subject to Hamilton's principle.[23] Varying Θ^{abcd} leads to (3.2.4). Since φ does not occur in $\mathcal{L}_S$, varying this field has the same effect as in **GR1**, and leads to the Klein-Gordon equation (3.2.3).[24]

If we want to exclude **SR2** from being background independent on this definition, we need to identify some means of demarcating actions such as that of Rosen as being pathological. To this end, Pooley suggests introducing a further criterion into the variational definition of

[23] For further discussion of this theory, see [125].

[24] One also needs to consider variations of g_{ab}. Rather than the EFEs, this leads to an equation that relates Θ^{abcd}, g_{ab} and φ [125, p. 696].

background independence, to arrive at the following: [104, p. 28]

Definition 8. (Background independence, variational) *A theory is* background independent *iff its solution space is determined by a generally-covariant action, (i) all of whose dependent variables are subject to Hamilton's principle, and (ii) all of whose dependent variables represent physical fields.*

The idea is that **SR2** fails to satisfy (ii) because, in Rosen's action, the field Θ^{abcd} is unphysical (as Pooley states, Θ^{abcd} "makes no impact on the evolution of g_{ab} and φ and hence, were it a genuine element of reality, it would be completely unobservable" [104, p. 29]). Though this seems to overcome the problem in this case, it is unclear whether other *further* cases—such as theories involving 'clock fields'—can be accommodated in a parallel manner [104, pp. 29-32].[25] To see what is meant here, turn now to the programme of *parameterisation*.[26]

Consider a simple case of parameterisation, in which one starts with the (Lorentz covariant) action for a theory such as **SR1**, defined with respect to inertial frame coordinates. To construct the parameterised theory corresponding to **SR1**, one treats the four coordinate fields X^{μ} of this formulation as themselves dependent variables ('clock fields'), writes them as functions of arbitrary coordinates, $X^{\mathsf{a}} = X^{\mathsf{a}}(x^{\nu})$,[27] and re-expresses the Lagrangian in terms of these new variables. Hamilton's principle is applied to the original dynamical variables, now conceived of as functions of x^{ν}, *and* to the clock fields X^{a}.

As Pitts states [96, p. 8-9], a parameterised special relativistic theory will have clock fields in the Poincaré invariant combination

$$\eta_{\mu\nu} := \eta_{\mathsf{ab}} X^{\mathsf{a}}{}_{,\mu} X^{\mathsf{b}}{}_{,\nu}, \tag{3.4.4}$$

[25] The example of a theory with clock fields is also mentioned at [95, p. 359].

[26] Discussed by e.g. Kuchař [70, 71].

[27] Having rewritten indices $\mu \to \mathsf{a}$ ($\mathsf{a} = 0 \ldots 3$) to indicate this elevation in status. The significance of this notational shift shall become clear in §4.2, where a parallel is drawn between clock fields and *tetrad fields*.

through which the metric field $\eta_{\mu\nu}$ is *defined* in terms of the X^{a}; $\eta_{\mathsf{ab}} = \mathrm{diag}\,(-1, 1, 1, 1)$ (so that η_{ab} is a *confined object* in the sense of §3.3.2); and the comma denotes differentiation with respect to arbitrary coordinates x^{ν}. In the case of **SR1**, stationarity under variations of φ leads to an equation for φ and X^{a} that is satisfied just in case φ satisfies the Lorentz-covariant Klein-Gordon equation with respect to the X^{a}. Stationarity under variations of the X^{a} yields equations that are automatically satisfied if the first equation is satisfied [104, pp. 29-30].[28]

In such a theory, do the X^{a} count as 'physical fields'? If they *do* so count in the case of e.g. the parameterised version of **SR1**, then there is a worry that Def. 8 delivers the incorrect verdict that such a theory *is* background independent – when intuitively it is *not*, since it is a reformulation of **SR1**. To resolve this matter, first note with Pooley that, unlike Θ^{abcd} above, the clock fields "certainly encode something physical, since they encode the metrical facts ... But there is also a sense in which they do not themselves directly represent something physical: coordinate systems are not physical objects" [104, p. 31].[29]

While it is true that clock fields, on the above presentation, are *constructed* from a given coordinate system, the relevance of this observation is, ultimately, unclear. The reason for this is that, in the resulting parameterised theory—KPMs of which may be specified to be tuples $\langle M, X^{\mathsf{a}}, \Phi\rangle$—the fields X^{a} manifestly *do* encode physical facts, insofar as they replace the metric field of the original theory. Thus, there do exist good reasons to take the clock fields to be physical fields, in spite of the means via which they are constructed.

A different question is the following: do the X^{a} qualify as *fixed fields*? Regardless of the theory from which we began (in our example, **SR1**), the answer appears to be *no*, for

[28] In fact, this result holds more generally [95, p. 9].

[29] Note also Pooley's observation that encoding a flat metric via clock fields *à la* (3.4.4) does not uniquely determine those clock fields [104, p. 31], for if X^{a} corresponds to one such set of fields, then so will any set X'^{a} where the X'^{a} are related to the X^{a} by a Lorentz transformation. This means that the X^{a} contain redundancy; 'internal' Lorentz transformations $X^{\mathsf{a}} \to X'^{\mathsf{a}}$ should be regarded as mere gauge redescriptions. In our view, the fact that the X^{a} exhibit gauge redundancy does not preclude their being physical fields – or at least, does not preclude some of their degrees of freedom from being counted as physical (cf. e.g. the vector potential of electromagnetism). Thus, the relevance of Pooley's observation here is unclear (cf. §4.2).

the X^{a} are subject to a variational principle. However, this *need not be so*, for consider Pitts' observation that, since stationarity under variations of the X^{a} yields equations that are automatically satisfied if equations of motion for the other fields are satisfied, "it does not much matter whether X^{a} are varied or not" [96, p. 9]. If we *do not* vary the X^{a}, then such fields violate Def. 8 of background independence.

The upshot, then, is the following: if the X^{a} *are* varied, then parameterised versions of theories such as **SR1** appear to pose a problem for Def. 8, for such theories, being reformulations of intuitively background dependent theories, are themselves intuitively background dependent. On the other hand, there is room to argue that the X^{a} *should not* be varied – in which case, such parameterised theories *do* come out as background dependent, on Def. 8. To overcome this contingent dependence on whether the clock fields of a parameterised theory are varied or not, one option is to fix (i) of Def. 8 to read: 'all of whose dependent variables *must* be subject to Hamilton's principle, in order to recover the dynamical laws of the theory.'

Finally, let us change tack, by remarking that Def. 8 should be considered only a *sufficient* condition for background independence, for in principle it is possible to construct theories consisting entirely of dynamical fields, with no corresponding action principle (i.e. no Lagrangian associated to the theory). Such theories may qualify, intuitively, as background independent, yet due to the absence of an appropriate action, would not so count on Def. 8.

One illustration of something like this theme arises through consideration of GR with asymptotic boundary conditions – for example, with the condition that the metric field g_{ab} be flat at spatial infinity (cf. §2.2). There exist (at least) three options available for the advocate of Def. 8 in assessing the background independence of this 'theory'. First, she may claim that a theory plus boundary conditions is not itself a new theory, and that background independence of a theory plus boundary conditions (these fixing a $\mathcal{B} \subset \mathcal{K}$) should be assessed with respect to $\mathcal{D} \subset \mathcal{K}$ (recall Fig. 2.1). In this case, since GR comes out as background independent on Def. 8, so too does GR with asymptotic boundary conditions.

A second option here is to consider the specified class $\mathcal{B}$ of BPMs of GR, in this case picked out by the condition of asymptotic flatness, to define the KPMs of a new theory:[30] GR plus these asymptotic boundary conditions. Then, background independence should be assessed with respect to whether the dynamical laws which pick out $\mathcal{D} \cap \mathcal{B} \subset \mathcal{B}$ can be obtained as per Def. 8. Since, however, in this theory we restrict at the level of KPMs to those models satisfying the specified boundary conditions, to derive the laws picking out $\mathcal{D} \cap \mathcal{B}$ from an action principle just is to derive the laws picking out $\mathcal{D}$ from an action principle. On this interpretation—according to which $\mathcal{B}$ picks out the KPMs of a new theory—a theory such as GR with asymptotic boundary conditions can also be background independent, on Def. 8.

A third position is the following. Once again, consider e.g. GR plus asymptotic boundary conditions to be a new theory; KPMs are picked out by $\mathcal{K}$; DPMs are now picked out by $\mathcal{D} \cap \mathcal{B}$ – so that in this case boundary conditions are imposed *dynamically*, whereas in the second possibility discussed above, they were imposed at the level of *kinematics*. In this case, in order for the theory to satisfy Def. 8, one would need to derive from an action principle dynamical equations which pick out such a $\mathcal{D} \cap \mathcal{B}$. It is not clear, however, that this can in general be achieved – that is, it is not clear that one can in general derive boundary conditions from an action principle. But if this is the case, then it looks like such theories are *not* background independent, on Def. 8 and this third approach – unless, as mentioned above, one regards Def. 8 merely as a *sufficient* condition on background independence.

3.5 Belot's Proposal

In the above, we have seen definitions of background independence in terms of absolute objects (Def. 5); fixed fields (Def. 6); and variational principles (Def. 8) – all three face difficulties when presented with various problem cases. In this section, we turn to a fourth account of

[30] This possibility was raised in §2.2.

background independence, due to Belot; the central 'morals' of this approach are claimed to be the following: [14, pp. 2872-2873]

1. Background (in)dependence comes in degrees: some theories are background (in)dependent, others nearly so, and others fall somewhere in between.
2. A theory can fail to be fully background independent in virtue of asymptotic boundary conditions.
3. The extent of the background (in)dependence of a theory is not a strictly formal one: in particular, it depends on how one thinks of the geometric structure of each solution and on what sorts differences between solutions one takes to be unphysical.

In §§3.5.1, 3.5.2, and 3.5.3, we recall how Belot arrives at these results; then, in §3.5.4, we raise some potential problems for Belot's programme.

3.5.1 Geometrical and Physical Degrees of Freedom

Central to Belot's account are the notions of *geometrical* and *physical* degrees of freedom. Beginning with the former, consider a KPM of a given theory; Belot tells us that we must be able to identity some *geometrical structure* in that model; moreover, there may exist a rule for when the geometrical structures of two KPMs are *equivalent*. Then, "we can look at the set of spacetime geometries that arise in [models] of the theory and at the set of equivalence classes of such geometries under the relation of geometrical equivalence ... We call any set of variables that parameterise this latter set the *geometrical degrees of freedom* of the theory relative to the geometrisation" [14, p. 2874].

To put this slightly differently, Belot's story regarding geometrical degrees of freedom runs as follows. First, for a theory $\mathcal{T}$, take a KPM $\mathcal{M} = \langle M, O_1, \ldots, O_n \rangle$, and identify certain objects in this model as 'geometrical structure'.[31] Marking such structure with a superscript G, we can write this KPM as $\mathcal{M} = \langle M, O^G, O_1 \ldots, O_{n-1} \rangle$. Now consider the geometrical structures O^G and O'^G associated respectively to two models $\mathcal{M}$ and $\mathcal{M}'$ of $\mathcal{T}$. In interpreting the models of $\mathcal{T}$, we may sometimes wish to regard O^G and O'^G as *equivalent* geometrical structure. For example, in models of SR, if $O^G = O'^G = \eta_{ab}$, then clearly these geometrical structures are equivalent, for they are the very same object. Similarly, in GR, we may wish to regard $O^G = g_{ab}$ and $O'^G = d^* g_{ab}$, $d \in \mathrm{Diff}\,(M)$, as *equivalent* geometrical structures.

Now consider the set $\mathcal{G}$ of the geometrical structures in all DPMs of $\mathcal{T}$. In accordance with the above rule, we can identify subsets of $\mathcal{G}$ which correspond to equivalence classes of geometrical structures. According to Belot's proposal, we should consider the *reduced* set of geometrical objects of the theory, call it $\tilde{\mathcal{G}}$, which is constructed from $\mathcal{G}$ by *quotienting* by these classes of equivalent geometries (so that each element of $\tilde{\mathcal{G}}$ corresponds to one such class in $\mathcal{G}$). The degrees of freedom needed to parameterise this latter set $\tilde{\mathcal{G}}$ are what Belot calls the *geometrical degrees of freedom* of $\mathcal{T}$.

With the above in mind, now introduce Belot's notion of physical degrees of freedom. Consider again the space $\mathcal{K}$ of KPMs of a theory $\mathcal{T}$. Often, we wish to interpret two models as representing the same possible world; the models are then said to be gauge-equivalent – see §2.2. In the presence of gauge-related models, we may quotient the space $\mathcal{K}$ of $\mathcal{T}$ to arrive at a reduced space $\tilde{\mathcal{K}}$, elements of which correspond to equivalence classes of gauge-related models in $\mathcal{K}$; the thought is that elements of $\tilde{\mathcal{K}}$ contain no redundant 'gauge' degrees of freedom. On Belot's account, the degrees of freedom needed to parameterise $\tilde{\mathcal{D}}$—the gauge-quotiented class of *DPMs* of $\mathcal{T}$—are the theory's *physical degrees of freedom*.[32]

Belot's position on background independence can then be put as follows: a theory is back-

[31]For clarity of exposition, we assume that there is just one such object representing 'geometrical structure' in the model.

[32]The reasons for which physical degrees of freedom are taken to parameterise $\tilde{\mathcal{D}}$ rather than $\tilde{\mathcal{K}}$ are discussed in §3.5.3.

ground independent just if its number of geometrical degrees of freedom coincides with its number of physical degrees of freedom (see Def. 10). The upshot of this definition is that, in a background independent theory, to each model $\tilde{\mathcal{M}} \in \tilde{\mathcal{D}}$, there corresponds *unique* geometrical structure $\tilde{O}^G \in \tilde{\mathcal{G}}$, in the sense of previous subsection. Thus, the geometrical structure $\tilde{O}^G$ acts as a *passport* for the associated model $\tilde{\mathcal{M}}$. In the following subsections, we see how this definition plays out in less abstract terms.

3.5.2 Background Independence

With the above notions of geometrical and physical degrees of freedom in hand, Belot defines four 'degrees' of background independence, as follows: [14, pp. 2877-2878]

Definition 9. (Full background dependence, Belot) *A field theory is* fully background dependent *if it has no geometrical degrees of freedom: every solution is assigned the same spacetime geometry as every other solution.*

Definition 10. (Full background independence, Belot) *A field theory is* fully background independent *if all of its physical degrees of freedom correspond to geometrical degrees of freedom: two solutions correspond to the same physical geometry iff they are gauge equivalent.*

Definition 11. (Near background dependence, Belot) *A field theory is* nearly background dependent *if it has only finitely many geometrical degrees of freedom: quotienting the space of geometries that arise in solutions of the theory by the relation of geometrical equivalence yields a finite-dimensional space.*

Definition 12. (Near background independence, Belot) *A field theory is* nearly background independent *if it has a finite number of non-geometrical degrees of freedom: there is some N such that for any geometry arising in a solution of the theory, the space of gauge equivalence classes of solutions with that geometry is no more than N-dimensional.*

In effect, these definitions measure the degree of background independence by looking at how many of the physical degrees of freedom correspond to geometrical degrees of freedom. The limiting cases are full background dependence (in which there are no geometrical degrees of freedom) and full background independence (in which all physical degrees of freedom are geometrical degrees of freedom); intermediate cases are also possible [14, p. 2878].

3.5.3 Applications

It is instructive to consider the application of the above definitions to certain case studies. Beginning with **SR1**, Belot writes of this theory that "the space of [KPMs] is the space of twice-differentiable φ; and the equation of motion is the Klein-Gordon equation, [(3.2.1)]. Here we have full background-dependence: the Minkowski metric η_{ab} is [*sic*] provides the fixed stage against which the scalar field evolves" [14, p. 2868]. The reason for this background dependence is that "each solution has the geometry of Minkowski spacetime, so there are no geometrical degrees of freedom" [14, p. 2878].

On **GR1**, Belot's account also seems to deliver the correct result: "the theory is ... fully background-independent: two solutions share the same geometry if and only if they are related by a diffeomorphism if and only if they are gauge equivalent" [14, p. 2878]. The expectation is that to each physical degree of freedom there corresponds a distinct geometrical degree of freedom, though whether there is indeed a one-one correspondence here is unclear – we return in §3.6 to assess whether this claim is strictly correct.

Now consider **SR2**. In **SR2**, the flatness of the spacetime metric is not fixed *ab initio* by its being specified as a fixed field; rather, it is determined dynamically by (3.2.4). This means that generic KPMs may be parameterised by non-trivial geometrical degrees of freedom. However, in the *DPMs* of the theory, the metric field g_{ab} is constrained to be flat, so there are no geometrical degrees of freedom.[33] Having made this move, Belot's account again issues the intuitively correct verdict that **SR2** is (fully) background dependent.

Finally, it is worth considering one advertised advantage of Belot's account: that it permits judgements of *intermediate* background independence. To see this, consider the sector of GR in which solutions are asymptotically flat at spatial infinity. Here, $M \cong \mathbb{R}^4$, and the only field is a Lorentzian metric field g_{ab} required to be twice-differentiable, globally hyperbolic, to approximate Euclidean geometry in a suitable sense at spatial infinity,[34] and to satisfy (2.1.1).[35] On this theory, Belot writes:

> The most natural thing to say is that this theory lies between paradigmatic non-background-independent theories like those in which fields propagate against the backdrop of Minkowski spacetime and paradigmatic background-independent theories like spatially compact general relativity. On the one hand, there are no fields on spacetime, fixed or dynamical, that encode a fixed background structure such as a geometry—indeed, locally the field of the theory has all of the freedom of the metric field of ordinary spatially compact general relativity. On the other, there is also a sense in which the boundary conditions of the theory ensure that any solution has the structure of Minkowski spacetime at infinity—and this is reflected in the fact that the theory is not generally covariant. [14, p. 2871]

Again, Belot's account offers an intuitively plausible judgement on this example: The theory is nearly background independent, but the fixing of the metric behaviour at infinity prevents it from being fully so. To see this, note that in this theory, two solutions correspond to the same geometry if and only they are related by a diffeomorphism;[36] for Belot, two solutions

[33] This illustrates why, when assessing whether a theory is background independent *à la* Belot, one must do so with respect to DPMs, rather than KPMs.

[34] This fixes $\mathcal{B} \subset \mathcal{K}$ in GR. In this section, however, for simplicity we consider this $\mathcal{B}$ to fix the *KPMs* of the theory of asymptotically flat GR, as per the discussion in §2.1.

[35] This fixes the space of DPMs $\mathcal{D} \subset \mathcal{K}$ of the theory.

[36] Recall our comment on GR in §3.5.1.

are gauge equivalent if and only if they are related by a diffeomorphism asymptotic to the identity at spatial infinity.[37] Then, "if one quotients the family of solutions sharing a given geometry by the relation of gauge equivalence, the result is a ten-dimensional space" [14, pp. 2878-2879] – accordingly, the theory satisfies Def. 12.[38]

As an aside, it is interesting to consider what one will make of asymptotically flat GR on Belot's definition if one does *not* endorse Belot's view that models related by a diffeomorphism asymptotic to the identity at spatial infinity are not to be considered gauge-equivalent, but rather embraces the interpretational account, according to which such models *are* gauge-equivalent (recall §2.2). In this case, gauge-equivalent models are those models related by diffeomorphisms, including diffeomorphisms at infinity, so there exists a one-one correspondence between physical and geometrical degrees of freedom, so that the theory comes out as fully background *independent*.

3.5.4 Problems for Belot's Programme

Belot's account appears to deliver the correct verdict on the background independence of a number of examples. Nevertheless, there are problems. In this subsection, we first present some methodological concerns, before turning to apparent counterexamples.

[37] Crucially, Belot's understanding of symmetry-related models as not always being gauge equivalent (see §2.2) is at play here.

[38] Another advertised advent of Belot's account—as pointed to in §3.5—is that it allows us to make sense of theories which are *ambiguously* background independent (see [14, §§3.3, 7.1]). Note also that Belot's account delivers the correct verdict in other cases, such as that of GR plus Jones-Geroch 'dust' (see §3.3.3).

The Priority of Geometry

Recall Brown's statement—made on the basis of prior discussions due to Anderson [5, p. 342]—that[39]

> Nothing in the form of the equations [of GR] *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, but, significantly, whose signature is indeterminate. ... The 'chronogeometric' or 'chronometric' significance of g_{ab} is not given a priori. [19, p. 160]

Such a position regarding the geometric import of the fields defined on M in a spacetime theory such as GR is clearly at odds with Belot's approach, which employs a prior specification of geometrical objects in any model of the theory in question – with background independence assessed with respect to these 'geometrical' objects. Though this is not a fatal problem for Belot's account, it does highlight that there is more to identifying 'geometry' in spacetime theories than Belot is apt to admit. Moreover, if one *does* endorse such a Brown-Anderson line, then one may be pushed to provide an account of background independence which treats all fields on a par (such as Def. 6 or Def. 8), rather than one involving any such prior demarcation of geometry.

Apparent Counterexamples

Perhaps more significantly, there also exist cases in which Belot's programme does not deliver the intuitively correct results – in part because Belot's definitions (presented in §3.5.2) are insufficiently closely tied to whether the fields of the theory in question are fixed or dynamical;

[39] Notation in this quotation has been amended to render it consistent with this thesis; there is no change in content. Whether such rewriting is completely innocuous is questionable in the light of [140]; nevertheless, we here set such worries aside.

and in part due to the fact that even after addressing this issue, geometrical degrees of freedom *need not* be in one-one correspondence with physical degrees of freedom in an intuitively background independent theory. Let us consider these two potential problems in turn.

First, consider a spacetime theory with KPMs $\langle M, g_{ab}, \varphi \rangle$, where φ is a real scalar field, subject to the specification that in *every* model, the field g_{ab} adopts an *inequivalent* configuration, and in each case such inequivalent configurations are in one-one correspondence with configurations of the field φ, which are *also* inequivalent across every possible model. In that case, *even though* the g_{ab} field is not subject to dynamical evolution and intuitively qualifies as 'background', the physical degrees of freedom of the theory correspond to the geometrical degrees of freedom – so Def. 10 is satisfied.

Indeed, even if the geometrical structures of the theory in question *are* coupled to the non-geometrical structures via the equations of motion, Belot's account faces difficulties. For example, for a general relativistic theory in which g_{ab} is coupled to the Φ via (2.1.1) and the dynamical equations for the Φ, there may in principle exist many distinct configurations of the Φ for the *same* g_{ab}. To illustrate, consider coupled Einstein-Maxwell theory (i.e. GR coupled to electromagnetism), KPMs of which are $\langle M, g_{ab}, F_{ab} \rangle$, where F_{ab} is the Faraday tensor. DPMs of this theory are picked out by the Einstein field equations (2.1.1), and the source-free Maxwell equations[40]

$$F^{ab}{}_{;b} = 0, \tag{3.5.1}$$

$$F_{[ab;c]} = 0. \tag{3.5.2}$$

Written in the notation of differential forms, (3.5.1) and (3.5.2) read, respectively

[40]These equations are constructed via the well-known *minimal coupling* scheme – for discussion, see [21].

$$dF = 0, \tag{3.5.3}$$

$$d * F = 0, \tag{3.5.4}$$

where $*$ is the Hodge operator. Taken together, (3.5.3) and (3.5.4) are formally invariant under

$$F \longleftrightarrow *F. \tag{3.5.5}$$

At the level of electric and magnetic fields,[41] such a transformation can be written

$$E_i \longleftrightarrow -B_i \tag{3.5.6}$$

$$B_i \longleftrightarrow E_i. \tag{3.5.7}$$

This is an example of *electromagnetic duality*.[42] Recalling that the stress-energy tensor of electromagnetism reads

$$T^{ab} = F^{ac}F^{b}_{\ \ c} - \frac{1}{4}g^{ab}F_{cd}F^{cd}, \tag{3.5.8}$$

one may verify that (3.5.8) is *invariant* under (3.5.5). Thus, if $\langle M, g_{ab}, F_{ab}\rangle$ is a DPM of this theory with stress-energy tensor T^{ab} associated to F_{ab}, then so too is $\langle M, g_{ab}, (*F)_{ab}\rangle$, with inequivalent Faraday tensor, but *identical* stress-energy tensor. Thus, even in GR, there exist

[41] Encoded in F_{ab} – see e.g. [63, p. 556].

[42] For recent philosophical discussion, see [101, 108].

scenarios in which different configurations of matter fields correspond to the same geometry – thereby violating Belot's definition of background independence.

In the first of the above cases, we have a situation in which Belot's criterion for background independence is satisfied; nevertheless, the theory does not intuitively count as background independent. In the second case, we have a situation in which Belot's criterion for background independence is not satisfied in an intuitively background independent theory.

3.6 Close

In this chapter, we have discussed *general covariance* (§3.1), *diffeomorphism invariance* (§3.2), and *absolute objects* (§3.3). Such background in hand, we then presented four definitions of background independence:

- Def. 5, in terms of the absence of absolute objects in a theory (§3.3.1).
- Def. 6, in terms of the absence of a theory formulation involving fixed fields (§3.4.1).
- Def. 8, in terms of variational principles (§3.4.2).
- Def. 10, in terms of matching geometrical and physical degrees of freedom (§3.5.2).

All four of these definitions face problems. In the following chapter, we embark upon an in-depth application of the definitions to four classical spacetime theories. Our goal, in doing so, is not only to provide a more concrete appraisal of the background independence of these theories than has hitherto been achieved; but also to continue to assess the relative merits of the above definitions, and the extent to which such problems may be overcome.

Chapter 4

Classical Spacetime Theories

In this chapter, we evaluate four classical spacetime theories—Newtonian gravitation (§4.1.2), Newton-Cartan theory (§4.1.3), teleparallel gravity (§4.2.2), and Kaluza-Klein theory (§4.3)—with respect to the question of their background independence. In each case, we present a comprehensive application and discussion of the definitions of background independence presented in the previous chapter; many important points regarding both these definitions, and the theories to which they are applied, emerge therefrom.

4.1 Newtonian Gravity and Newton-Cartan Theory

In the field formulation of Newtonian gravitation (NG),[1] gravity is mediated by a potential; in the presence of a (non-constant) gravitational potential, the matter distribution will change.[2] NG is theoretically equivalent to another theory,[3] known as *Newton-Cartan theory* (NCT). In

[1] Developed by Laplace and Poisson [67, p. 265].

[2] See §4.1.2. In fact, the precise sense in which this is so is underdetermined in NG – see footnote 17.

[3] In the sense of [141–143], i.e. that there exists an isomorphism between $\tilde{\mathcal{K}}_1$ of $\mathcal{T}_1$ and $\tilde{\mathcal{K}}_2$ of $\mathcal{T}_2$ such that the empirical predictions corresponding to $\tilde{\mathcal{D}}_1 \subset \tilde{\mathcal{K}}_1$ are identical to the empirical predictions corresponding

NCT, by contrast with NG, gravity is 'geometrised' in an approach similar to that of GR: the geometrical properties of spacetime depend on the distribution of matter, and, conversely, gravitational effects are manifestations of this geometry.

In this section, we first introduce in §4.1.1 the formal notion of a *classical spacetime*, this being the setting for both NG and NCT. These theories we then discuss, respectively, in §§4.1.2 and 4.1.3.[4] Finally, in §4.1.4, we engage in an in-depth assessment of the background independence of both of these theories.

4.1.1 Classical Spacetimes

Define a *classical spacetime* to be a quadruple $\langle M, t_{ab}, h^{ab}, \nabla_a \rangle$, where M is a smooth four-dimensional differentiable manifold; t_{ab} is a smooth, symmetric tensor field on M of signature $(1,0,0,0)$; h^{ab} is a smooth, symmetric tensor field on M of signature $(0,1,1,1)$; ∇_a is a derivative operator on M;[5] and the following three conditions hold:

$$h^{ab}t_{ab} = 0, \tag{4.1.1}$$

$$\nabla_a t_{bc} = 0, \tag{4.1.2}$$

$$\nabla_a h^{bc} = 0. \tag{4.1.3}$$

We refer to (4.1.1) as an *orthogonality* condition, and (4.1.2) and (4.1.3) as *compatibility* conditions.

$\tilde{\mathcal{D}}_2 \subset \tilde{\mathcal{K}}_2$.

[4]Much of our presentation follows [34, 46, 78].

[5]For the formal definition of such a derivative operator, see e.g. [78, p. 49].

By the 'signature' of t_{ab} and h^{ab} is meant the following. The signature condition for t_{ab} is the requirement that, at every $p \in M$, the tangent space has a set of four basis vectors $\zeta_{\mathsf{a}}{}^a$ ($\mathsf{a} = 0, \ldots, 3$)[6] such that $t_{ab}\zeta_{\mathsf{a}}{}^a\zeta_{\mathsf{b}}{}^b = 0$ if $\mathsf{a} \neq \mathsf{b}$, and[7]

$$t_{ab}\zeta_{\mathsf{a}}{}^a\zeta_{\mathsf{a}}{}^b = \begin{cases} 1 & \text{if } \mathsf{a} = 0 \\ 0 & \text{if } \mathsf{a} = 1, 2, 3. \end{cases} \tag{4.1.4}$$

Importantly, note that t_{ab} does not satisfy the non-degeneracy condition to be a metric – for example, if the vectors $\zeta_{\mathsf{a}}{}^a$ are as above at some point $p \in M$, then $t_{ab}\zeta_1{}^a = 0$ at p, even though $\zeta_1{}^a \neq 0$ [78, p. 250]. The signature condition for h^{ab}, similarly, is the requirement that, at every point, the cotangent space has a basis $\sigma^{\mathsf{a}}{}_a$ ($\mathsf{a} = 0, \ldots, 3$) such that $h^{ab}\sigma^{\mathsf{a}}{}_a\sigma^{\mathsf{b}}{}_b = 0$ if $\mathsf{a} \neq \mathsf{b}$, and

$$h^{ab}\sigma^{\mathsf{a}}{}_a\sigma^{\mathsf{a}}{}_b = \begin{cases} 0 & \text{if } \mathsf{a} = 0 \\ 1 & \text{if } \mathsf{a} = 1, 2, 3. \end{cases} \tag{4.1.5}$$

Note that h^{ab} is not the inverse of a metric, i.e. there is no h_{ab} such that $h_{ab}h^{bc} = \delta^c{}_a$.[8]

Let us now consider the interpretation of the t_{ab}, h^{ab} and ∇_a in a classical spacetime. Beginning with t_{ab}, this can be considered a 'temporal metric' – though, as we have seen, strictly it does not satisfy the metric non-degeneracy condition. Given any vector θ^a at a point $p \in M$, we can take its 'temporal length' to be $\left(t_{ab}\theta^a\theta^b\right)^{1/2}$. We further classify θ^a

[6] Thus far, sans serif indices are simply labels for basis vectors, though they acquire a greater significance in §4.2.2. Note also that labelling of sans serif indices begins at 0 by convention; the reason for this shall become apparent in §4.2.2.

[7] Here, no summation convention over sans serif indices is deployed.

[8] To see why, suppose that the $\sigma^{\mathsf{a}}{}_a$ are as above at some $p \in M$, so that $h^{ab}\sigma^0{}_a = 0$. Then, if there were a tensor h_{ab} at that point such that $h_{ab}h^{bc} = \delta^c{}_a$, it would follow that $0 = h_{ab}h^{bc}\sigma^0{}_c = \delta^c{}_a\sigma^0{}_c = \sigma^0{}_a$, contradicting the assumption that the $\sigma^{\mathsf{a}}{}_a$ form a basis for the cotangent space at p [78, p. 250].

as either *timelike* or *spacelike*, depending on whether its temporal length is positive or zero, respectively. It follows from the signature of t_{ab} that the subspace of spacelike vectors at any point is three-dimensional [78, p. 250]. At any $p \in M$, we can find a covector t_a, unique up to sign, such that $t_{ab} = t_a t_b$ [78, pp. 250-251]. We say that a classical spacetime is *temporally orientable* if there exists a continuous, *globally-defined* covector field t_a that satisfies this condition at every $p \in M$. Any such field t_a is called a *temporal orientation*. A timelike vector θ^a is *future-directed* relative to t_a if $t_a\theta^a > 0$; otherwise it is *past-directed* [78, p. 251].[9]

In the following, we restrict to temporally orientable classical spacetimes. Since in this case t_{ab} can be constructed from t_a, we accordingly denote classical spacetimes which satisfy the temporal orientability condition by $\langle M, t_a, h^{ab}, \nabla_a \rangle$.[10] For such spacetimes, (4.1.1) and (4.1.2) can be written, respectively, [78, p. 251]

$$h^{ab} t_b = 0, \tag{4.1.6}$$

$$\nabla_a t_b = 0. \tag{4.1.7}$$

From (4.1.7), it follows that t_a is *closed*, i.e. $\nabla_{[a} t_{b]} = 0$. Hence, at least locally, it must be *exact*, i.e. of the form $t_a = \nabla_a t$, for some smooth function t. We call any such function t a *time function* [78, p. 251]. Call a hypersurface $S \subset M$ *spacelike* if, at all $p \in S$, all vectors tangent to S are spacelike; this is equivalent to the requirement that all time functions be constant on S.[11] We can think of spacelike hypersurfaces as 'simultaneity slices'. If M is

[9] We understand a smooth curve to be *timelike* (respectively *spacelike*) if its tangent vectors are of this character at every point along the curve. Similarly, a timelike curve is understood to be *future-directed* (respectively *past-directed*) if its tangent vectors are so at every point along the curve. [78, p. 251]

[10] If a classical spacetime admits one temporal orientation t_a, then it also admits a second temporal orientation, $-t_a$ [78, p. 251]. In cases in which we restrict attention to classical spacetimes that are temporally orientable, we by convention select the temporal orientation with positive sign.

[11] A time function t is constant on S iff, given any vector θ^a tangent to S at some $p \in S$, $\theta^a \nabla_a t = 0$. But $t_a\theta^a = \theta^a \nabla_a t$, so the latter condition holds iff all vectors tangent to S are spacelike. [78, p. 252]

simply connected, then there exists a globally defined time function [78, p. 252]. In that case, spacetime can be decomposed into a one-parameter family of global simultaneity slices.

Now consider h^{ab}. Informally, this can be considered a 'spatial metric', though the details of how h^{ab} assigns lengths to vectors are subtle. We cannot take the spatial length of a vector θ^a to be $\left(h_{ab}\theta^a\theta^b\right)^{1/2}$, because the latter is not well-defined (as we have seen, there does not exist an h_{ab} such that $h^{ab}h_{bc} = \delta^a{}_c$). If, however, θ^a is *spacelike*, then we can use h^{ab} to assign a spatial length to it indirectly. The key proposition is the following: [78, p. 253]

Proposition 1. *Let $\langle M, t_a, h^{ab}, \nabla_a\rangle$ be a classical spacetime. Then the following conditions hold at all points $p \in M$:*

1. *For all σ_a, $h^{ab}\sigma_b = 0$ iff σ_a is a multiple of t_a.*
2. *For all μ^a, μ^a is spacelike iff there is a σ_a such that $h^{ab}\sigma_b = \mu^a$.*
3. *For all σ_a and σ'_a, if $h^{ab}\sigma_b = h^{ab}\sigma'_b$, then $h^{ab}\sigma_a\sigma_b = h^{ab}\sigma'_a\sigma'_b$.*

Proof: See [78, pp. 253-254]. □

Using Prop. 1, we can assign a *spatial length* to a spacelike vector θ^a by taking this to be $\left(h^{ab}\sigma_a\sigma_b\right)^{1/2}$, where σ_b is a vector such that $h^{ab}\sigma_b = \theta^a$ [78, p. 254]. (2) guarantees the existence of such a σ_a; (3) guarantees the irrelevance of the particular choice of σ_a [78, p. 254]. Finally, note that because h^{ab} is not invertible, we cannot raise *and* lower indices with it. We can, however, *raise* indices: for example, if $R^a{}_{bcd}$ is the Riemann curvature tensor associated with ∇_a (see below), then we understand $R^{ab}{}_{cd}$ to be the field $h^{be}R^a{}_{ecd}$ [78, p. 256].

Now consider the derivative operator ∇_a in a classical spacetime $\langle M, t_a, h^{ab}, \nabla_a\rangle$. We first present the following general result regarding derivative operators on M: [78, p. 51]

Proposition 2. *Let ∇_a and ∇'_a be derivative operators on M. Then there exists a smooth* *symmetric tensor field $C^{a}{}_{bc}$ on M satisfying, for all smooth tensor fields $\alpha^{a_1\dots a_r}_{b_1\dots b_s}$ on M:*

$$\left(\nabla'_c - \nabla_c\right)\alpha^{a_1\dots a_r}_{b_1\dots b_s} = \alpha^{a_1\dots a_r}_{db_2\dots b_s}C^{d}{}_{cb_1} + \dots + \alpha^{a_1\dots a_r}_{b_1\dots b_{s-1}d}C^{d}{}_{cb_s} - \alpha^{da_2\dots a_r}_{b_1\dots b_s}C^{a_1}{}_{cd} - \dots - \alpha^{a_1\dots a_{r-1}d}_{b_1\dots b_s}C^{a_r}{}_{cd} \tag{4.1.8}$$

Conversely, given any derivative operator ∇_a on M and any smooth symmetric tensor field $C^{a}{}_{bc}$ on M, if ∇'_a is defined by (4.1.8), then ∇'_a is also a derivative operator on M.

Proof: See [78, pp. 51-52]. □

In what follows, if ∇'_a and ∇_a are derivative operators on M that, together with the field $C^{b}{}_{cd}$, satisfy (4.1.8), then we shall, following Malament [78, p. 53], write $\nabla'_a = \left(\nabla_a, C^{b}{}_{cd}\right)$. Now, for a classical spacetime, the conditions (4.1.2) and (4.1.7) do not determine a *unique* ∇_a. In fact, we have the following result: [78, p. 256]

Proposition 3. *Let $\langle M, t_a, h^{ab}, \nabla_a\rangle$ be a classical spacetime. Let $\nabla'_a = \left(\nabla_a, C^{b}{}_{cd}\right)$ be a new derivative operator on M. Then ∇'_a is compatible with t_a and h^{ab} iff $C^{a}{}_{bc}$ is of the form*

$$C^{a}{}_{bc} = 2h^{ad}t_{(b}\kappa_{c)d}, \tag{4.1.9}$$

where κ_{ab} is a smooth, antisymmetric tensor field on M.

Proof: See [78, pp. 256-257]. □

It is also worth considering some properties of the Riemann tensor $R^{a}{}_{bcd}$ associated to ∇_a in a classical spacetime. As always (see e.g. [78, p. 69]), $R^{a}{}_{bcd}$ satisfies the algebraic relations $R^{a}{}_{b(cd)} = R^{a}{}_{[bcd]} = 0$. The conditions (4.1.2) and (4.1.7) further imply that $t_a R^{a}{}_{bcd} =$

$R^{(ab)}{}_{cd} = 0$ [78, p. 258]. It follows therefrom that if we raise all indices with h^{ab}, the resulting field R^{abcd} satisfies $R^{ab(cd)} = R^{a[bcd]} = R^{(ab)cd} = 0$; this in turn implies that $R^{abcd} = R^{cdab}$ [78, p. 258]. As usual, the Ricci tensor R_{ab} is defined as $R_{ab} = R^{c}{}_{abc}$. If, for the derivative operator ∇_a associated to a classical spacetime, $R^{a}{}_{bcd} = 0$ at some $p \in M$, then we say that spacetime is *flat* at p.[12] Similarly, we say that a classical spacetime is *spatially flat* at some $p \in M$ if $R^{abcd} = 0$ at p. This is motivated by the following proposition: [78, p. 261]

Proposition 4. (Spatial flatness) *Let $\langle M, t_{ab}, h^{ab}, \nabla_a \rangle$ be a classical spacetime. Then the following conditions are equivalent at every point $p \in M$:*

1. $R^{abcd} = 0$.
2. $R^{ab} = 0$.
3. $R_{ab} = t_{(a}\varphi_{b)}$ *for some* φ_a.

Furthermore, given any spacelike hypersurface $S \subset M$, these conditions hold throughout S iff parallel transport of spacelike vectors within S is, at least locally, path independent.

Proof: See [78, pp. 261-262]. □

Prop. 4 will prove important in our discussions of NCT in §4.1.3. It is worth noting that intermediate between the curvature conditions $R^{a}{}_{bcd} = 0$ and $R^{abcd} = 0$ is the condition $R^{ab}{}_{cd} = 0$. This holds throughout M iff parallel transport of spacelike vectors along arbitrary curves is, at least locally, path independent [78, p. 263].[13]

To close this subsection, let us make some comments on the physical interpretation of classical spacetimes. Following Malament [78, p. 252], we introduce the following interpretive

[12] Since this condition holds iff parallel transport of vectors on M relative to ∇_a is, at least locally, path-independent – violations of which constitute the *definition* of spacetime curvature.

[13] Here we still restrict attention to spacelike vectors, but consider their transport along arbitrary curves in M, rather than just curves confined to a particular spacelike hypersurface $S \subset M$.

principles.[14] For all smooth curves $\gamma : I \to M$:[15]

- γ is timelike 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.

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. If the particle has a mass m, then its four-acceleration field satisfies

$$F^a = m\xi^b \nabla_b \xi^a, \tag{4.1.10}$$

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 NG and NCT.

4.1.2 Newtonian Gravity

In the four-dimensional, field-theoretic formulation of NG, models are specified as tuples $\langle M, t_a, h^{ab}, \nabla_a, \varphi, \Phi \rangle$. The first four elements of this tuple represent classical spacetime structure, as per §4.1.1. In addition, however, two new elements are introduced in NG: a real scalar

[14] If these principles are taken as *fundamental postulates*, then the advocate of the *dynamical approach* to spacetime theories (developed by Brown [19]), according to which spacetime structure is a *codification* of certain dynamical properties of matter fields, may demur. Any tension between the following postulates and the dynamical approach is, however, illusory, for the advocate of the dynamical approach may simply regard such points as non-fundamental facts, which can be given some further dynamical explanation.

[15] I is an open interval in $\mathbb{R}$.

field φ, which represents the *gravitational potential*, and Φ, which is a placeholder for the matter fields of the theory. In NG, one imposes flatness of ∇_a via the field equation[16]

$$R^a{}_{bcd} = 0. \tag{4.1.11}$$

A second field equation of NG is *Poisson's equation*,

$$h^{ab}\nabla_a\nabla_b\varphi = 4\pi\rho. \tag{4.1.12}$$

The mass density function ρ is associated to the Φ in a manner parallel to the relationship between the stress-energy tensor T_{ab} and the matter fields in relativistic spacetime theories.[17] Finally, the gravitational force on a point particle of mass m is given by $-mh^{ab}\nabla_b\varphi$. It follows from (4.1.10) that if the particle is subject to no forces except gravity, and if it has four-velocity

[16]On this way of imposing the flatness of ∇_a in NG, the theory is akin to **SR2** – in which flatness of the spacetime metric is imposed dynamically via (3.2.4) [104, p. 18]. As we have seen, there exists a theory, **SR1**, corresponding to **SR2**, in which the spacetime metric is fixed to be the Minkowski metric η_{ab} in all KPMs of the theory (i.e. kinematically, rather than dynamically); in **SR1**, unlike **SR2**, the spacetime metric is a *fixed field*. Given existence of such a theory **SR1** corresponding to **SR2**, it should be possible to construct an analogous theory to the above formulation of NG, in which ∇_a is fixed to be flat in this sense. In either case, however, we would expect such fixing of the flatness of ∇_a in NG to mean that there exists background structure in the theory; to this issue we return in §4.1.4. Relatedly, one might question why the **SR2**-type formulation of NG should be taken as the default presentation, rather than the **SR1**-type formulation. The most straightforward answer to this question is that the former presentation renders most explicit the parallels between NG and NCT, discussed in §§4.1.3 and 4.1.4.

[17]Two points are in order here. First, just as it is a mistake in GR to conflate the matter fields Φ and their associated stress-energy tensor T_{ab} (cf. footnote 7 in §2.1), it is a mistake to conflate the matter fields Φ of NG and their associated mass density function ρ. This leads to the second point: The mass density function ρ in NG, as with the stress-energy tensor of GR, affords only a *partial* description of the associated matter fields Φ (cf. §3.5.4). For instance, ρ omits information on the *velocity field* of matter in the theory. (We might then say that while in GR T_{ab} is *partial* and *non-fundamental* (since it may be given a closed-form expression in terms of the matter fields of the theory), in NG ρ is *partial* and *fundamental*.) As Dewar observes, this particular omission leads to a pernicious (and, until recently, largely unnoticed) indeterminism in NG – see [32, p. 4].

ξ^a, then it satisfies

$$-\nabla^a \varphi = \xi^b \nabla_b \xi^a. \tag{4.1.13}$$

4.1.3 Newton-Cartan Theory

In NCT, the gravitational potential φ in NG is absorbed into the derivative operator ∇_a, which is accordingly rendered partly dynamical, and hence generically curved.[18] Models of NCT are therefore specified by tuples $\langle M, t_a, h^{ab}, \nabla_a, \Phi \rangle$. The conversion from models of NG to models of NCT proceeds via Trautman's *geometrisation theorem* [132]:[19]

Theorem 1. (Geometrisation theorem) *Let $\langle M, t_a, h^{ab}, \nabla_a \rangle$ be a classical spacetime with ∇_a flat ($R^a{}_{bcd} = 0$). Further, let φ and ρ be smooth real-valued functions on M satisfying Poisson's equation $h^{ab}\nabla_a\nabla_b\varphi = 4\pi\rho$. Finally, let $\hat{\nabla} = (\nabla_a, C^a{}_{bc})$, with $C^a{}_{bc} = -t_b t_c \nabla^a \varphi$. Then all of the following hold:*

1. *$\langle M, t_a, h^{ab}, \hat{\nabla} \rangle$ is a classical spacetime.*
2. *$\hat{\nabla}$ is the unique derivative operator on M such that, for all timelike curves on M with four-velocity field ξ^a,*

$$\xi^b \hat{\nabla}_b \xi^a = 0 \quad \Longleftrightarrow \quad \xi^b \nabla_b \xi^a = -\nabla^a \varphi \tag{4.1.14}$$

3. *The curvature field $\hat{R}^a{}_{bcd}$ associated with $\hat{\nabla}$ satisfies:*

[18] We say *partly*, for the NCT curvature conditions (4.1.16) and (4.1.17) (discussed below), as well as the compatibility conditions (4.1.2) and (4.1.7), imply that not *all* degrees of freedom associated to the derivative operator ∇_a of NCT are subject to dynamical evolution.

[19] This version due to Malament [78, pp. 267-268].

$$\hat{R}_{bc} = 4\pi\rho t_b t_c \tag{4.1.15}$$

$$\hat{R}^{a\ c}_{\ b\ d} = \hat{R}^{c\ a}_{\ d\ b} \tag{4.1.16}$$

$$\hat{R}^{ab}{}_{cd} = 0 \tag{4.1.17}$$

Proof: See [78, pp. 268-269]. □

It follows from (4.1.14) that particles subject to a gravitational force in NG—and thereby exhibiting geodesic deviation with respect to ∇_a—exhibit geodesic motion with respect to $\hat{\nabla}_a$. (4.1.15) is the geometrised version of Poission's equation; (4.1.16) holds throughout M iff parallel transport of spacelike vectors in M is, at least locally, path independent; (4.1.17) 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 ($\nabla^{[a}\xi^{b]} = 0$) [78, p. 281].[20] We see from Prop. 4 that (4.1.15) implies spatial flatness. As Malament writes of this latter result: [78, p. 262]

> This is striking. It is absolutely fundamental to the idea of geometrised Newtonian theory that *spacetime* is curved (and gravitational is just a manifestation of that curvature). Yet the basic field equation of the theory itself rules out the possibility that *space* is curved.

With these points in mind, let us now consider the reverse translation, from NCT to NG. This proceeds via Trautman's *recovery theorem* (presented in [132]):[21]

[20] There exists an alternative formulation of NCT due to Künzle [72, 73] and Ehlers [39], in which (4.1.17) is dropped. The motivation for doing so is that this appears to be necessary in order for NCT to qualify as a limiting version of GR [78, pp. 270-271]. In this thesis, we focus on the formulation of NCT due to Trautman [132], which includes (4.1.17), since in this case a full translation between NG and NCT is possible. In the Künzle-Ehlers version of NCT, only a translation between NCT and a *generalised* form of NG is attainable. (In this generalised version of NG, the gravitational force acting on a particle of unit mass is given by a vector field, but it need not be of the form $\nabla^a \varphi$; in addition, field equations involve a 'rotation field', ω_{ab} [78, pp. 270-271].)

[21] This version due to Malament [78, pp. 274-275].

Theorem 2. (Recovery theorem) *Let $\langle M, t_a, h^{ab}, \nabla_a \rangle$ be a classical spacetime satisfying:*

$$R_{ab} = 4\pi\rho t_a t_b \tag{4.1.18}$$

$$R^{a\ c}_{\ b\ d} = R^{c\ a}_{\ d\ b} \tag{4.1.19}$$

$$R^{ab}{}_{cd} = 0 \tag{4.1.20}$$

for some smooth scalar field ρ on M. Then given any point $p \in M$, there is an open set O containing p, a smooth scalar field φ on O, and a derivative operator $\tilde{\nabla}$ on O such that all the following hold on O:

1. *$\tilde{\nabla}_a$ is compatible with t_a and h^{ab}.*

2. *$\tilde{\nabla}_a$ is flat.*

3. *For all timelike curves with four-velocity fields ξ^a,*

$$\xi^b \nabla_b \xi^a = 0 \quad \Longleftrightarrow \quad \xi^b \tilde{\nabla}_b \xi^a = -\tilde{\nabla}^a \varphi. \tag{4.1.21}$$

4. *$\tilde{\nabla}_a$ satisfies Poisson's equation $h^{ab}\tilde{\nabla}_a \tilde{\nabla}_b \varphi = 4\pi\rho$.*

The pair $\langle \tilde{\nabla}, \varphi \rangle$ is not unique. A second pair $\langle \tilde{\nabla}', \varphi' \rangle$, defined on the same open set O, will satisfy the stated conditions iff:

A. $\nabla^a \nabla^b (\varphi' - \varphi) = 0$

B. $\tilde{\nabla}' = \left(\tilde{\nabla}, C'^a{}_{bc}\right)$, *where* $C'^a{}_{bc} = t_b t_c \nabla^a (\varphi' - \varphi)$.

Proof: See [78, pp. 275-277]. □

Thm. 2 tells us that if ∇_a arises as the geometrisation of the pair $\langle \tilde{\nabla}_a, \varphi \rangle$, then, for any field ψ such that $\nabla^a \nabla^b \psi = 0$, it also arises as the geometrisation of the pair $\langle \tilde{\nabla}'_a, \varphi' \rangle$, where $\varphi' = \varphi + \psi$, and $\tilde{\nabla}'_a = \left(\tilde{\nabla}_a, t_b t_c \nabla^d \psi \right)$. As Malament points out [78, p. 278], this observation captures the fact that, in NG, the gravitational force on a particle is determined only up to a factor of $m \nabla^a \psi$, where $\nabla^a \psi$ is constant on any one spacelike hypersurface, but can change over time. We return to this observation in §4.1.4.

4.1.4 Background Independence

Newtonian Gravitation

So much for the formal differences between NG and NCT. What, then, of their background independence? To begin, consider ∇_a in NG. As presented in §4.1.2, flatness of ∇_a in NG is imposed dynamically, via (4.1.11). It would, however, also be possible to impose flatness at the level of KPMs, by restricting such models to those in which ∇_a is flat.[22] Though a naïve approach to background independence[23] might deliver the verdict that ∇_a counts as background structure in this 'fixed field' formulation of NG, but not in the formulation presented in §4.1.2, more sophisticated accounts of background independence such as Def. 5 and Def. 6 deliver the correct verdict that ∇_a counts as background structure in either case.[24]

With this in mind, turn now in earnest to whether NG is background independent. This theory is sometimes cited as a paragon of background dependence; and, indeed, it is easy to see

[22] Cf. footnote 16.

[23] Such as the equation of this notion with the absence of fixed fields – see §3.2.4.

[24] Relatedly, consider the possibility of an alternative formulation of NG (or, for that matter, NCT), in which the conditions (4.1.1)-(4.1.3) are imposed *dynamically*, via field equations. Doing so has specific unattractive consequences, for (unlike e.g. g_{ab} in **SR2**), there exists no natural 'physical' interpretation of KPMs where, e.g., the compatibility of the spatial and temporal metrics is violated. In turn, this raises the question of whether there even exists an analogue of **SR2** for these theories.

why: both h^{ab} and t_a are *fixed fields* which, in a Lagrangian formulation of NG, would not be subject to Hamilton's principle. Due to the presence of such fields, NG violates Def. 6; due to their non-variational status, NG also violates Def. 8; since such fields qualify as Andersonian absolute objects, NG violates Def. 5.

The existence of such fixed fields leads to there being multiple distinct configurations of the fields $\langle \varphi, \Phi \rangle$ corresponding to the same classical spacetime $\langle M, t_a, h^{ab}, \nabla_a \rangle$ – which we take to represent *geometrical structure* in the models of NG.[25,26] Thus, in Belot's language (§3.5.1), the number of physical degrees of freedom of the theory exceeds the number of geometrical degrees of freedom, so NG does not satisfy Def. 10.

While it is straightforward to issue the above verdict on NG, the situation regarding Belot's account is, in fact, rather delicate. First, one must consider whether NG satisfies Belot's Def. 9, of *full* background dependence. Recall from Prop. 3 that the derivative operator ∇_a of NG is not unique; thus, there exist models of the theory with the same $\langle \varphi, \Phi \rangle$, but set in classical spacetimes differing in their derivative operators. In more detail: for a given DPM of NG $\langle M, t_{ab}, h^{ab}, \nabla_a, \varphi, \Phi \rangle$, $\langle M, t_{ab}, h^{ab}, \nabla'_a, \varphi, \Phi \rangle$ is also a DPM of NG, where $\nabla'_a = \left(\nabla_a, C^b{}_{cd} \right)$, $C^a{}_{bc} = 2h^{ad} t_{(b} \kappa_{c)d}$, and $\nabla_{[a} \kappa_{b]c} = 0$.[27,28] A schematic response to this result is the following: 'The geometries of such models (are/are not) equivalent, and the models [do/do not] exhibit gauge redundancy'. Since, however, it is not plausible to understand such models as having equivalent geometries yet as not being gauge equivalent (since these models are identical in all other respects), we turn now to the other three possibilities.

First, suppose that the geometries of such models *are* equivalent, and the models *are* con-

[25] On the form of NG presented above, while t_a and h^{ab} are fixed in all KPMs, ∇_a is equivalent up to isomorphism across DPMs (up to transformations of the form discussed below) – cf. footnote 13 of §3.2.4.

[26] Below, we modify our assumption that $\langle M, t_a, h^{ab}, \nabla_a \rangle$ constitutes geometrical structure in a model of NG.

[27] One uses (4.1.8) to show that (4.1.12) is invariant under this change of derivative operator; to show that (4.1.11) is also invariant, it is convenient to use $R'^a{}_{bcd} = R^a{}_{bcd} + 2\nabla_{[c} C^a{}_{d]b} + 2C^e{}_{b[c} C^a{}_{d]e}$ (see [78, p. 70]), where $R'^a{}_{bcd}$ is the Riemann tensor associated to ∇'_a.

[28] Imposing flatness of ∇_a at the level of KPMs (as discussed above) does not dissolve this multiplicity, unless one restricts at the level of KPMs to *one* such choice of ∇_a.

sidered gauge equivalent. In that case, in this regard, to each physical configuration, there corresponds a unique geometrical configuration – so such a redundancy does not inhibit NG's being background independent on Belot's account.[29] Second, suppose that the geometries of such models are *not* equivalent, and the models are *not* considered gauge equivalent. Then, it is still the case that (in this regard) each distinct model corresponds to a unique geometry, so again, the background independence of NG on Belot's account is not threatened.[30] Third, consider the possibility that such geometries are *not* equivalent, but the models in question *are* gauge equivalent.[31] In that case, Def. 10 is straightforwardly not satisfied. Thus, on this latter view, there exists an additional sense in which NG violates Def. 10.

On the second and third interpretive routes above, that the geometrical objects of the theory (and in particular ∇_a) may be inequivalent across DPMs means that it is *not* the case that there are *no* geometrical degrees of freedom in the theory. Therefore, it appears on such interpretations that one *cannot* regard NG as *fully* background dependent, on Def. 9.

Indeed, a similar point arises out of a second (more familiar) multiplicity in the models of NG. To see this, recall from Thm. 2 that for a given model with derivative operator and gravitational potential $\langle \nabla_a, \varphi \rangle$,[32] there exists a class of other models of NG with $\langle \nabla'_a, \varphi' \rangle$—where $\varphi' = \varphi + \psi$; $\nabla'_a = \left(\nabla_a, t_b t_c \nabla^d \psi\right)$; and ψ is a field such that $\nabla^a \nabla^b \psi = 0$—which correspond to the same model of NCT. This being so, it is argued in e.g. [68, 78, 141] that it is plausible to understand these as being gauge equivalent.[33]

[29] Though recall that the background independence of NG according to Def. 10 fails for other reasons – namely, that h^{ab} and t_a are fixed across all models.

[30] Though see footnote 29.

[31] Such a view may be defensible on the interpretational approach to symmetries, but, absent some story regarding *how* such models with inequivalent geometries can be gauge equivalent, is not immediately defensible on the motivational approach (cf. §2.2). Note also that if one takes 'equivalent geometries' to mean *physically* equivalent geometries, then it is questionable whether two gauge-equivalent models can possess inequivalent geometries – in which case, this third option is excluded.

[32] Having selected such a derivative operator, as per the above discussion.

[33] On this multiplicity of models in NG, Malament points out at [78, p. 279] that if we are dealing with a bounded mass distribution—i.e. if ρ has compact support on every spacelike hypersurface—and moreover if we require that the gravitational field die off as one approaches spatial infinity, then the multiplicity disappears. The reason for this is that if $\nabla^a \psi$ is constant on spacelike hypersurfaces *and* goes to zero at spatial infinity, then it

In this case, we face the same four interpretive options as above. First, to regard such classes of models as having equivalent geometries yet not being gauge equivalent is now a live option, since the φ field now differs from model to model within such classes. As above, one may also argue that such models have inequivalent geometries, but *are* gauge equivalent. In either case, the biconditional of Def. 10 is violated, so such an interpretation again yields a further sense in which NG is not background independent, on Belot's definition. Alternatively, if one regards all such models as having equivalent geometries and as being gauge equivalent (*à la* [78, p. 278], [141, §6]); or as having inequivalent geometries and as not being gauge equivalent (*à la* [14, p. 4]), then this multiplicity of models does not give rise to further road-blocks to NG being background independent, on Def. 10. Again, however, any interpretation here on which the derivative operators are *not* equivalent across models results in it *not* being the case that there are *no* geometrical degrees of freedom in NG – thereby posing problems for any attempt to understanding this theory as fully background dependent, as per Def. 9.

The intricacies regarding the interplay between NG and Belot's account do not end here. Consider now Knox's suggestion that "the spacetime structure of NG is represented not by the flat ... connection usually made explicit in formulations, but by the sum of the flat connection and the gravitational field" [68, p. 863]. If one takes this seriously, then one must understand the geometrical structure in models of NG to be picked out not by $\langle M, t_{ab}, h^{ab}, \nabla_a \rangle$, but by $\langle M, t_{ab}, h^{ab}, \nabla_a, \varphi \rangle$. Given such a tuple, a unique ρ is picked out via (4.1.12); moreover, given a different tuple $\langle M, t_{ab}, h^{ab}, {\nabla'}_a, \varphi' \rangle$, the *same* ρ is picked out, via the above reasoning. If one understands such models to have equivalent geometries,[34] and to be gauge-equivalent,[35] then it is clear that NG *does* satisfy Def. 10.[36]

must vanish everywhere. Thus, by imposing such boundary conditions, this multiplicity collapses, and cannot further threaten the background independence of NG according to Def. 10.

[34] Which is plausible, if, with Knox, one considers only the *sum* of contributions from a given $\langle {\nabla'}_a, \varphi' \rangle$ to be physical [67, p. 870].

[35] Which, as mentioned above, is a mainstream position.

[36] Throughout this section, we set aside further complications regarding the 'pernicious indeterminism' of footnote 17.

This result is important, for intuitively it is the incorrect verdict – the reason being that in *every* model of NG, h^{ab} and t_b *remain fixed*. The problem stems from the fact that Belot's programme does not adequately account for the possibility of *multiple* geometric objects in a theory being counted as geometrical structure in the sense of §3.5.1.[37] To overcome this problem, a caveat must be added to Belot's account: If multiple geometric objects in the theory are considered to be geometrical structure, they must *all* be inequivalent in all non-gauge equivalent models of the theory. Thus, we see that NG presents a novel and concrete problem case for Belot's account, which can be used to tune the programme.

Our verdict on the background independence of NG is, then, as follows: While the theory violates Def. 5, Def. 6, and Def. 8, a number of subtleties inhibit a clear-cut conclusion on NG's satisfaction of Def. 9. These results in hand, let us close by turning to whether NG is diffeomorphism invariant on Def. 3. The straightforward answer here is *no*, for, treating h^{ab} as a fixed field and performing an arbitrary smooth coordinate transformation on (4.1.12), one extracts the condition for the associated transformation matrix $M^a{}_b$,

$$M^a{}_c M^b{}_d h^{cd} = h^{ab}. \tag{4.1.22}$$

Newton-Cartan Theory

This discussion of NG in hand, turn now to NCT. Since in this theory, gravitation is 'geometrised' in a manner similar to GR, one might expect NCT also to be background independent. This, however, is *not* the case, for both h^{ab} and t_a *remain fixed*. While ∇_a is rendered *partly* dynamical,[38] such fixing of the geometrical objects of NCT renders the theory back-

[37]Below, we find the same result in the case of NCT. This is not surprising, for classes of models of NG differing in $\langle \nabla_a, \varphi \rangle$ correspond to the *same* model of NCT (cf. §4.1.3).

[38]But still not *fully*, in light of curvature conditions (4.1.16) and (4.1.17) and metric compatibility (4.1.2) and (4.1.7) – cf. footnote 18.

ground dependent, on Def. 5, Def. 6, and Def. 8.

It might, in addition, be tempting to state that such fixing of the geometrical objects of NCT renders the theory background dependent on Belot's account. This, however, is not correct. Recall that for full background independence, Belot requires a one-one match between the geometrical degrees of freedom of the theory—which here parameterise the difference between tuples $\langle t_b, h^{ab}, \nabla_a \rangle$ across models—and the physical degrees of freedom—which here parameterise the difference between tuples $\langle t_b, h^{ab}, \nabla_a, \Phi \rangle$ across gauge inequivalent models. However,[39] since the Ricci tensor associated to the derivative operator ∇_a and the mass density function ρ of the matter fields Φ are coupled via the field equation (4.1.15), on Belot's account it turns out that there *does* exist a unique geometry corresponding to each distinct physical configuration – meaning that NCT *does* satisfy Def. 10.

As in the case of NG, the above is an intuitively incorrect result. However, by introducing the above-mentioned caveat that, if multiple geometric objects in the theory are considered to be geometrical structure, they must *all* be inequivalent in all non-gauge equivalent models of the theory, one may circumvent this verdict.

Let us close with two comments. First, the non-uniqueness of the derivative operator ∇_a in a classical spacetime (c.f. §4.1.1) may—depending on one's interpretation—threaten the background independence of NCT in exactly the same manner as in the case of NG, discussed above. Second, unlike NG, NCT *is* diffeomorphism invariant on Def. 3, for the dynamical laws of the theory (4.1.15)-(4.1.17) are invariant under arbitrary smooth coordinate transformations.

[39]Modulo concerns analogous to those regarding GR, as presented in §3.5.4. Cf. also footnote 36.

4.2 General Relativity and Teleparallel Gravity

NCT is the geometrised theory of gravity associated with the force theory NG. The theoretical equivalence[40] between NG and NCT raises the question: Does there exist a force theory associated to GR? As Knox states, it appears that the answer here is *yes*, the theory in question being *teleparallel gravity* (TPG) [67, §2]. In TPG, the Levi-Civita connection of GR, which has curvature but no torsion, is replaced with a *Weitzenböck connection*, with torsion but no curvature. In this section, we review the details of GR (§4.2.1) and TPG (§4.2.2). We then present explicit translation theorems relating the two theories, in the style of §4.1.3. This done, we assess in §4.2.3 the background independence of TPG. Given the analogy between NG and NCT, and TPG and GR, it is natural to ask: Is the pattern of judgements regarding the background independence of NG and NCT preserved in the case of TPG and GR?

4.2.1 General Relativity

Let us recall the essential details of GR.[41] First, define a *relativistic spacetime* to be a pair $\langle M, g_{ab} \rangle$, where M is a four-dimensional differentiable manifold, and g_{ab} is a Lorentzian metric field on M. 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 $T^a_{\ bc} X^b Y^c = \nabla_b X^b Y^a - X^a \nabla_b Y^b - [X, Y]^a$, vanishes—and (ii) metric compatible – in the sense that $\nabla_a g_{bc} = 0$. Given the uniqueness of this derivative operator, ∇_a is *not* included in models of GR, which are denoted by triples $\langle M, g_{ab}, \Phi \rangle$, where Φ is again a placeholder for the matter fields of the theory.[42] The dynamical equations of GR are the Einstein field equations, (2.1.1), plus dynamical equations associated to the matter fields.[43]

[40] See footnote 3.

[41] See also §2.1.

[42] Clearly, the existence of such a unique derivative operator is in contrast with NG and NCT, for which we saw in §§4.1.2 and 4.1.3 that no *unique* such operator exists.

[43] Cf. footnote 8 of §2.1.

4.2.2 Teleparallel Gravity

Tetrad Fields

Knox claims that TPG is to GR what NG is to NCT [67, §2].[44] To present TPG, we must first define a *tetrad field*. Suppose that at some $p \in M$, we are working within the constraints of a coordinate basis for the tangent space T_pM at p. Generically, such coordinates will not be orthogonal. Nevertheless, at each $p \in M$ we can always find an orthonormal basis for T_pM: for every $p \in M$, T_pM has a set of four basis vectors ${h_\mathsf{a}}^a$ ($\mathsf{a} = 1 \ldots 4$) such that

$$g_{ab}{h_\mathsf{a}}^a{h_\mathsf{a}}^b = \begin{cases} +1 & \text{if } \mathsf{a} = 1 \\ -1 & \text{if } \mathsf{a} = 2, 3, 4 \end{cases} \tag{4.2.1}$$

and $g_{ab}{h_\mathsf{a}}^a{h_\mathsf{b}}^b = 0$ if $\mathsf{a} \neq \mathsf{b}$. In other words, at each $p \in M$, we can always find a set of four basis vectors ${h_\mathsf{a}}^a$ for T_pM such that

$$g_{ab}{h^a}_\mathsf{a}{h^b}_\mathsf{b} = \eta_\mathsf{ab}. \tag{4.2.2}$$

With this in mind, define a *tetrad field* ${h^a}_\mathsf{a}$ to be a set of four vector fields which at each $p \in M$ provide such an orthonormal basis for T_pM. In (4.2.2) and what follows, $(\mathsf{a}, \mathsf{b}, \ldots)$ are understood to be indices in an orthonormal basis provided by the tetrad field. Importantly, it follows as a lemma from (4.2.2) that[45]

[44] Although 'teleparallel gravity' can be used to refer to a family of theories using the Weitzenböck connection, the variant discussed here reproduces the results of GR exactly.

[45] To prove (4.2.3), consider $\eta_\mathsf{ab}{h^\mathsf{a}}_a{h^\mathsf{b}}_b{h^b}_\mathsf{c} = \eta_\mathsf{ab}{h^\mathsf{a}}_a{\delta^\mathsf{b}}_\mathsf{c} = \eta_\mathsf{ac}{h^\mathsf{a}}_a = h_{\mathsf{c}a} = g_{ab}{h^b}_\mathsf{c}$; cancelling the ${h^b}_\mathsf{c}$ yields the result. To prove (4.2.4), simply raise the b index in (4.2.3).

$$\eta_{\mathsf{ab}} h^{\mathsf{a}}{}_{a} h^{\mathsf{b}}{}_{b} = g_{ab} \tag{4.2.3}$$

$$h^{\mathsf{a}}{}_{a} h^{b}{}_{\mathsf{a}} = \delta^{b}{}_{a}. \tag{4.2.4}$$

The Weitzenböck Derivative and Teleparallel Gravity

Given a tetrad field, we define the components $\bar{\Gamma}^{\rho}{}_{\mu\nu}$ of the so-called *Weitzenböck derivative* $\bar{\nabla}_a$ to be [7, p. 3]

$$\bar{\Gamma}^{\rho}{}_{\mu\nu} := h^{\rho}{}_{\mathsf{a}} \partial_{\nu} h^{\mathsf{a}}{}_{\mu}. \tag{4.2.5}$$

$\bar{\nabla}_a$ is the derivative operator of TPG. This in hand, we declare that KPMs of TPG are triples $\langle M, h^{\mathsf{a}}{}_{a}, \Phi \rangle$, where $h^{\mathsf{a}}{}_{a}$ is a tetrad field on M, and, as before, Φ is a placeholder for the matter fields of the theory. Analogously with GR (and in contrast with NG and NCT), the Weitzenböck derivative $\bar{\nabla}_a$ is not included as a separate element in the model, for, given a tetrad field $h^{\mathsf{a}}{}_{a}$, there exists a unique such operator, defined via (4.2.5).

Non-Uniqueness of the Tetrad Fields

For any $p \in M$, the set of tetrad fields $h^{\mathsf{a}}{}_{a}$ for $T_p M$ is not unique. In fact, (4.2.2) only determines the tetrad field up to local Lorentz transformations [1, p. 13]. To see this, suppose that there exists another tetrad field $h'^{\mathsf{a}}{}_{a}$, also satisfying (4.2.2). Contracting both sides of

(4.2.3) within $h'_{\mathsf{c}}{}^{a} h'_{\mathsf{d}}{}^{b}$ and using (4.2.2) in terms of $h'^{\mathsf{a}}{}_{a}$, we obtain

$$\eta_{\mathsf{ab}} h'_{\mathsf{c}}{}^{a} h_{a}{}^{\mathsf{a}} h'_{\mathsf{d}}{}^{b} h_{b}{}^{\mathsf{b}} = \eta_{\mathsf{cd}}. \tag{4.2.6}$$

Defining $M_{\mathsf{a}}{}^{\mathsf{b}} := h'_{\mathsf{a}}{}^{a} h_{a}{}^{\mathsf{b}}$, we have

$$\eta_{\mathsf{ab}} M_{\mathsf{c}}{}^{\mathsf{a}} M_{\mathsf{d}}{}^{\mathsf{b}} = \eta_{\mathsf{cd}}, \tag{4.2.7}$$

so the $M_{\mathsf{a}}{}^{\mathsf{b}}$ satisfy the definition of Lorentz matrices. Rearranging the definition of $M_{\mathsf{a}}{}^{\mathsf{b}}$, we have

$$h'^{\mathsf{a}}{}_{a} = M^{\mathsf{a}}_{\mathsf{b}} h^{\mathsf{b}}{}_{a}. \tag{4.2.8}$$

Thus, any field $h'^{\mathsf{a}}{}_{a}$ which differs from a tetrad field $h^{\mathsf{a}}{}_{a}$ by a Lorentz transformation will also be a tetrad field. Considering geometrical objects of TPG such as $\bar{\Gamma}^{\rho}{}_{\mu\nu}$ and $\bar{T}^{\rho}{}_{\mu\nu}$,[46] it is then straightforward to verify that such objects are *invariant* under such Lorentz transformations of tetrad indices [1, p. 13].[47]

[46] $\bar{T}^{\rho}{}_{\mu\nu}$ is the torsion tensor (defined in §4.2.1) associated to $\bar{\nabla}_{a}$; this object should not be conflated with the stress-energy tensor of TPG, discussed below.

[47] $\bar{\Gamma}'^{\rho}{}_{\mu\nu} = h'^{\rho}{}_{\mathsf{a}} \partial_{\nu} h'^{\mathsf{a}}{}_{\mu} = \left(M^{T} M\right)_{\mathsf{a}}{}^{\mathsf{b}} h^{\rho}{}_{\mathsf{b}} \partial_{\nu} h^{\mathsf{a}}{}_{\mu} = \delta_{\mathsf{a}}{}^{\mathsf{b}} h^{\rho}{}_{\mathsf{b}} \partial_{\nu} h^{\mathsf{a}}{}_{\mu} = h^{\rho}{}_{\mathsf{a}} \partial_{\nu} h^{\mathsf{a}}{}_{\mu} = \Gamma^{\rho}{}_{\mu\nu}$. Since $\bar{T}^{\rho}{}_{\mu\nu}$ is defined directly from $\bar{\Gamma}^{\rho}{}_{\mu\nu}$, this object, and in turn the force and field equations of TPG ((4.2.21) and (4.2.29) – discussed in §4.2.2), are invariant under such transformations. Indeed, as stated at [1, p. 13], *all* spacetime-indexed quantities in TPG should be so invariant.

The Levi-Civita Derivative

In presentations of TPG, reference is sometimes made to the *Levi-Civita derivative*. The referent of this term should, however, be made explicit, for unlike GR, no Lorentzian metric g_{ab} is fundamental in TPG. To do so, first recall that, generically, the coefficients $\Gamma^{\rho}{}_{\mu\nu}$ of the Levi-Civita derivative are given by

$$2g_{\sigma\rho}\Gamma^{\rho}{}_{\mu\nu} = \partial_{\mu}g_{\sigma\mu} + \partial_{\mu}g_{\sigma\nu} - \partial_{\sigma}g_{\mu\nu}. \tag{4.2.9}$$

Using (4.2.3), the right hand side of (4.2.9) can be expressed in terms of the tetrad field as

$$\begin{aligned} 2g_{\sigma\rho}\Gamma^{\rho}{}_{\mu\nu} = \eta_{\mathsf{ab}}h^{\mathsf{b}}{}_{\mu}\partial_{\nu}h^{\mathsf{a}}{}_{\sigma} + \eta_{\mathsf{ab}}h^{\mathsf{a}}{}_{\sigma}\partial_{\nu}h^{\mathsf{b}}{}_{\mu} + \eta_{\mathsf{ab}}h^{\mathsf{b}}{}_{\nu}\partial_{\mu}h^{\mathsf{a}}{}_{\sigma} + \\ \eta_{\mathsf{ab}}h^{\mathsf{a}}{}_{\sigma}\partial_{\mu}h^{\mathsf{b}}{}_{\nu} - \eta_{\mathsf{ab}}h^{\mathsf{b}}{}_{\nu}\partial_{\sigma}h^{\mathsf{a}}{}_{\mu} - \eta_{\mathsf{ab}}h^{\mathsf{a}}{}_{\mu}\partial_{\sigma}h^{\mathsf{b}}{}_{\nu}. \end{aligned} \tag{4.2.10}$$

(4.2.10) can be considered a definition of the components of the Levi-Civita derivative in TPG in terms of the tetrad field.[48] Given this, it is useful to compute the difference of the components of the Weitzenböck and Levi-Civita derivatives; we find

$$\bar{\Gamma}^{\rho}{}_{\mu\nu} - \Gamma^{\rho}{}_{\mu\nu} = K^{\rho}{}_{\mu\nu}, \tag{4.2.11}$$

[48] Strictly, for this to be a definition one should also rewrite the $g_{\sigma\rho}$ on the left hand side of (4.2.10) in terms of the tetrad field, via (4.2.3).

where $K^{\rho}{}_{\mu\nu}$ is the so-called *contortion tensor* [7, p. 3], defined as

$$K^{a}{}_{bc} = \frac{1}{2}\left(\bar{T}_{b}{}^{a}{}_{c} + \bar{T}_{c}{}^{a}{}_{b} - \bar{T}^{a}{}_{bc}\right). \tag{4.2.12}$$

The Teleparallel Force Law

TPG is supposed to be "a gauge theory for the translation group" [7, p. 1]. In order to implement this, first consider a frame in which $h^{\mathsf{a}}{}_{\mu} = \partial_{\mu} x^{\mathsf{a}}$;[49] then augment the tetrad field with an extra term $B^{\mathsf{a}}{}_{\mu}$: [93, §3.1]

$$h^{\mathsf{a}}{}_{\mu} = \partial_{\mu} x^{\mathsf{a}} + B^{\mathsf{a}}{}_{\mu}. \tag{4.2.13}$$

Inspired by this decomposition, one then considers a translation of the tangent space coordinates, $x^{\mathsf{a}} \to x^{\mathsf{a}} + \Delta^{\mathsf{a}}$, under which, generically, we can write $B^{\mathsf{a}}{}_{\mu} \to B'^{\mathsf{a}}{}_{\mu}$. After such a transformation, the tetrad field becomes

$$h^{\mathsf{a}}{}_{\mu} \to h'^{\mathsf{a}}{}_{\mu} = \partial_{\mu} x^{\mathsf{a}} + \partial_{\mu}\Delta^{\mathsf{a}} + B'^{\mathsf{a}}{}_{\mu}. \tag{4.2.14}$$

Under what circumstances does $h'^{\mathsf{a}}{}_{\mu}$, as with $h^{\mathsf{a}}{}_{\mu}$, satisfy the definition of the tetrad field (4.2.2)? This result holds just if

$$B^{\mathsf{a}}{}_{\mu} \to B'^{\mathsf{a}}{}_{\mu} = B^{\mathsf{a}}{}_{\mu} - \partial_{\mu}\Delta^{\mathsf{a}}. \tag{4.2.15}$$

[49] This is possible locally in light of (3.4.4). Indeed, the analogue between (3.4.4) and (4.2.3) highlights the close connection between tetrad fields and clock fields.

In the formulation of TPG under consideration here, we *stipulate* that $B^{\mathsf{a}}{}_{\mu}$ has such transformation properties under a change $x^{\mathsf{a}} \to x^{\mathsf{a}} + \Delta^{\mathsf{a}}$; we then see that the tetrad field $h^{\mathsf{a}}{}_{\mu}$ is an *invariant object* under such transformations [1, p. 38].[50] Inspired by (4.2.15), it is natural to view TPG as an Abelian gauge theory of the translation group. With this in mind, we define an Abelian field strength tensor [7, p.2]

$$F^{\mathsf{a}}{}_{\mu\nu} := \partial_\mu B^{\mathsf{a}}{}_{\nu} - \partial_\nu B^{\mathsf{a}}{}_{\mu} \tag{4.2.16}$$

$$= \partial_\mu h^{\mathsf{a}}{}_{\nu} - \partial_\nu h^{\mathsf{a}}{}_{\mu}; \tag{4.2.17}$$

the second line here follows using (4.2.14). Now, in a coordinate basis, the definition of the torsion tensor (see §4.2.1) reads $T^{\lambda}{}_{\mu\nu} = \Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\lambda}{}_{\nu\mu}$. In the case of the Weitzenböck connection, we hence have

$$\begin{aligned} \bar{T}^{\lambda}{}_{\mu\nu} &= \bar{\Gamma}^{\lambda}{}_{\mu\nu} - \bar{\Gamma}^{\lambda}{}_{\nu\mu} \\ &= h_{\mathsf{a}}{}^{\lambda}\partial_\nu h^{\mathsf{a}}{}_{\mu} - h_{\mathsf{a}}{}^{\lambda}\partial_\mu h^{\mathsf{a}}{}_{\nu}. \end{aligned} \tag{4.2.18}$$

[50] In this language, $h^{\mathsf{a}}{}_{\mu}$ is *gauge-invariant* [1, p. 38], *pace* Knox [67, p. 273]. As a consequence, Knox's further claim that "the tetrad field is defined only up to a local gauge transformation, and this gauge freedom passes onto the connection defined in terms of the tetrads" [67, p. 273] is also incorrect. While it is possible that Knox has in mind here the non-uniqueness of the tetrad field discussed in §4.2.2, there also exist problems on this reading. First, while it may also be plausible to treat such Lorentz transformations of the tetrad fields as gauge transformations, it is confusing to run together these two possible senses of *gauge transformation* in TPG without qualification. Second, even putting this aside, the claim that "this gauge freedom passes onto the connection defined in terms of the tetrads" is incorrect, for we have seen that the components of the Weitzenböck derivative are formally *invariant* under such a transformation.

Multiplying through by $h^{\mathsf{b}}{}_{\lambda}$, it follows that

$$h^{\mathsf{a}}{}_{\lambda}\bar{T}^{\lambda}{}_{\mu\nu} = \partial_{\mu}h^{\mathsf{a}}{}_{\nu} - \partial_{\nu}h^{\mathsf{a}}{}_{\mu}, \tag{4.2.19}$$

wherefrom we obtain the result

$$F^{\mathsf{a}}{}_{\mu\nu} = h^{\mathsf{a}}{}_{\lambda}\bar{T}^{\lambda}{}_{\mu\nu}. \tag{4.2.20}$$

As in §4.1, defining a particle 4-velocity to be ξ^{a}, and recalling that the Lorentz force law[51] in covariant notation reads $\xi^{a}\nabla_{a}\xi^{b} = F^{b}{}_{a}\xi^{a}$,[52] we define by analogy the force law of TPG to be

$$\begin{aligned}\xi^{a}\bar{\nabla}_{a}\xi_{b} &= \xi_{\mathsf{a}}F^{\mathsf{a}}{}_{ba}\xi^{a}\\ &= \xi_{\mathsf{a}}h^{\mathsf{a}}{}_{c}\bar{T}^{c}{}_{ba}\xi^{a}\\ &= \xi^{c}\bar{T}_{cba}\xi^{a}.\end{aligned} \tag{4.2.21}$$

This equation predicts torsion-dependent deviations away from the geodesics of the Weitzenböck connection.

[51] Itself associated to the Abelian gauge theory of electromagnetism.

[52] Having set $q = m = 1$, where q is the charge of the particle in question and m is its mass.

Field Equations

Finally, we turn to the field equations of TPG – i.e., the analogue of the Einstein equations (2.1.1) in the theory. These equations—derivable from an action principle[53]—read: [7, p. 4]

$$\partial_\sigma \left(h S_{\mathsf{a}}{}^{\sigma\rho} \right) - 4\pi \left(h j_{\mathsf{a}}{}^{\rho} \right) = 0, \tag{4.2.22}$$

with[54]

$$h = \det \left(h^{\mathsf{a}}{}_{\mu} \right), \tag{4.2.23}$$

$$h j_{\mathsf{a}}{}^{\rho} = -\frac{1}{4\pi} h h_{\mathsf{a}}{}^{\lambda} S_{\mu}{}^{\nu\rho} \bar{T}^{\mu}{}_{\nu\lambda} + h_{\mathsf{a}}{}^{\rho} \mathcal{L}_G, \tag{4.2.24}$$

$$S^{\rho\mu\nu} = \frac{1}{2} \left(K^{\mu\nu\rho} - g^{\rho\nu} \bar{T}^{\sigma\mu}{}_{\sigma} + g^{\rho\mu} \bar{T}^{\sigma\nu}{}_{\sigma} \right), \tag{4.2.25}$$

$$\mathcal{L}_G = \frac{h}{16\pi} S^{\rho\mu\nu} \bar{T}_{\rho\mu\nu}. \tag{4.2.26}$$

Using the definition of the components of the Weitzenböck connection (4.2.5), we can write (4.2.22) as[55]

$$\partial_\sigma \left(h S_{\lambda}{}^{\sigma\rho} \right) - 4\pi \left(h t_{\lambda}{}^{\rho} \right) = 0, \tag{4.2.27}$$

[53] The relevant gravitational Lagrangian in TPG is $\mathcal{L}_G$, defined at (4.2.26).

[54] $h j_{\mathsf{a}}{}^{\rho}$ is related to gravitational energy-momentum in TPG – for details, see [7, p. 4].

[55] $t_{\lambda}{}^{\rho}$ is a *pseudotensor* encoding gravitational energy-momentum in TPG; there are thus clear and interesting parallels with the gravitational energy-momentum pseudotensor in GR (for some philosophical discussion of this object, see [55, 74, 97]).

with

$$h t_{\lambda}{}^{\rho} = \frac{h}{4\pi} \bar{\Gamma}^{\mu}{}_{\nu\lambda} S_{\mu}{}^{\nu\rho} + \delta_{\lambda}{}^{\rho} \mathcal{L}_G. \tag{4.2.28}$$

This is equivalent to the vacuum field equations of GR, $R_{ab} = 0$ [7, p. 5]. Analogously, in the presence of matter, the field equations of TPG become

$$\partial_{\sigma} \left(h S_{\lambda}{}^{\sigma\rho} \right) - 4\pi \left(h t_{\lambda}{}^{\rho} \right) = 4\pi \left(h T_{\lambda}{}^{\rho} \right). \tag{4.2.29}$$

This is equivalent to the Einstein field equations of GR with matter, (2.1.1) [7, p. 6].

Inter-Theory Translations

The above in hand, we are ready to prove geometrisation and recovery theorems, to translate from TPG to GR and vice versa (respectively). We begin with the geometrisation theorem.

Theorem 3. (Geometrisation Theorem, TPG) *Let $\bar{\mathcal{M}} = \langle M, h^{\mathsf{a}}{}_{\mu}, \Phi \rangle$ be a model of TPG, for which the field equations (4.2.29) hold, and let $\bar{\nabla}_a$ be the unique Weitzenböck derivative associated to $\bar{\mathcal{M}}$. Then all of the following hold:*

1. *There exists a unique metric field g_{ab} and associated Levi-Civita derivative ∇_a which can be constructed from $h^{\mathsf{a}}{}_{\mu}$ via (4.2.2) and (4.2.10), respectively.*
2. *∇_a is the unique derivative operator on M such that, for all timelike curves on M with four-velocity field ξ^a,*

$$\xi^b \nabla_b \xi^a = 0 \quad \Longleftrightarrow \quad \xi^a \bar{\nabla}_a \xi_b = \xi^c \bar{T}_{cba} \xi^a. \tag{4.2.30}$$

3. *The curvature field $R^a{}_{bcd}$ associated with ∇_a satisfies the Einstein field equations,* (2.1.1).

As a consequence, for each DPM $\bar{\mathcal{M}}$ of TPG, there exists a unique DPM $\mathcal{M}$ of GR.

Proof: (1) follows from (4.2.2) and (4.2.10). For (2), We begin with the force law of GR, $\xi^a\nabla_a\xi^b = 0$. Expanding the Levi-Civita derivative in a coordinate basis, we have

$$\xi^\mu\partial_\mu\xi^\nu + \xi^\mu\Gamma^\nu{}_{\rho\mu}\xi^\rho = 0. \tag{4.2.31}$$

Now using (4.2.11), we can write

$$\xi^\mu\partial_\mu\xi^\nu + \xi^\mu\bar{\Gamma}^\nu{}_{\rho\mu}\xi^\rho - \xi^\mu K^\nu{}_{\rho\mu}\xi^\rho = 0, \tag{4.2.32}$$

i.e.,

$$\xi^a\bar{\nabla}_a\xi^b = \xi^a K^b{}_{ca}\xi^c. \tag{4.2.33}$$

Using (4.2.12), we then have

$$\xi^a\bar{\nabla}_a\xi_b = \xi^c\bar{T}_{cba}\,\xi^a, \tag{4.2.34}$$

which is our desired result: the force law for TPG. The converse translation is also possible, so the result follows. For (3), recall from §4.2.2 that the dynamical equations of TPG (4.2.29) are equivalent to the Einstein equations (2.1.1), in terms of the curvature field $R^a{}_{bcd}$ associated to the Levi-Civita derivative ∇_a (see e.g. [1,7]). Thus, for any DPM of TPG, one can construct a unique Lorentizan metric field g_{ab} with associated Levi-Civita derivative, and those geometrical objects so constructed will satisfy (2.1.1). Thus, to every DPM $\bar{\mathcal{M}}$ of TPG, a unique DPM $\mathcal{M}$ of GR can be constructed therefrom. $\square$

This in hand, we turn now to the recovery theorem.

Theorem 4. (Recovery Theorem, TPG) *Let $\mathcal{M}$ be a model of GR satisfying the Einstein field equations* (2.1.1)*, and let ∇_a be the unique Levi-Civita connection associated to $\mathcal{M}$. Then all of the following hold:*

1. *There exists a tetrad field, $h^{\mathsf{a}}{}_a$, and unique associated Weitzenböck derivative $\bar{\nabla}_a$.*

2. *$\bar{\nabla}_a$ is flat (i.e. $\bar{R}^{\mu}{}_{\nu\rho\sigma} = 0$ for the associated Riemann tensor $\bar{R}^{\mu}{}_{\nu\rho\sigma}$).*

3. *For all timelike curves on M with four-velocity field ξ^a,*

$$\xi^b \nabla_b \xi^a = 0 \quad \Longleftrightarrow \quad \xi^a \bar{\nabla}_a \xi_b = \xi^c \bar{T}_{cba}\, \xi^a. \tag{4.2.35}$$

4. *The tetrad field $h^{\mathsf{a}}{}_a$ satisfies the field equations* (4.2.29).

The pair $\langle h^{\mathsf{a}}{}_a, \bar{\nabla}_a \rangle$ is not unique. A second pair $\langle h'^{\mathsf{a}}{}_a, \bar{\nabla}_a \rangle$ will also satisfy the stated conditions iff $h'^{\mathsf{a}}{}_a = M^{\mathsf{a}}{}_{\mathsf{b}} h^{\mathsf{b}}{}_a$, where $M^{\mathsf{a}}{}_{\mathsf{b}}$ is a Lorentz matrix (i.e. a matrix satisfying (4.2.7)*).*

As a consequence, for each DPM $\mathcal{M}$ of GR, there exists a class of DPMs $\bar{\mathcal{M}}$ of TPG, differing by Lorentz transformations of the orthonormal indices of the tetrad field.

Proof: For (1), a tetrad field $h^{\mathsf{a}}{}_a$ can be constructed from a Lorentzian metric g_{ab} via (4.2.4); an associated Weitzenböck derivative $\bar{\nabla}_a$ can then be constructed via (4.2.5). For (2), we need to prove that any such $\bar{\nabla}_a$ is flat, in the sense that the associated Riemann tensor $\bar{R}^{\mu}{}_{\nu\rho\sigma}$

vanishes. To do so, recall that in a coordinate basis, the Riemann tensor reads

$$R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} - \partial_\sigma \Gamma^{\mu}{}_{\nu\rho} + \Gamma^{\tau}{}_{\nu\sigma}\Gamma^{\mu}{}_{\tau\rho} - \Gamma^{\tau}{}_{\nu\rho}\Gamma^{\mu}{}_{\tau\sigma}. \tag{4.2.36}$$

Substituting in the definition of the components of the Weitzenböck derivative (4.2.5), we observe pairwise cancellation, yielding the result

$$\bar{R}^{a}{}_{bcd} = 0. \tag{4.2.37}$$

For (3), the result follows from the same calculations as those presented for (2) in Thm. 3. For (4), we again recall from §4.2.2 that the Einstein equations of GR (2.1.1) and the field equations of TPG (4.2.29) are inter-translatable, so given that $\mathcal{M}$ is a DPM of GR, the tetrad field $h^{\mathsf{a}}{}_{a}$ constructed therefrom will satisfy (4.2.29).

To prove non-uniqueness, first recall from §4.2.2 that if $h^{\mathsf{a}}{}_{a}$ satisfies (4.2.2), then so too does $h'^{\mathsf{a}}{}_{a} = M^{\mathsf{a}}{}_{\mathsf{b}} h^{\mathsf{b}}{}_{a}$. Thus, if a tetrad field $h^{\mathsf{a}}{}_{a}$ can be constructed from a Loretnzian metric field g_{ab}, then so also can a distinct tetrad field $h'^{\mathsf{a}}{}_{a}$. As with $h^{\mathsf{a}}{}_{a}$, we can construct a Weitzenböck derivative associated to $h'^{\mathsf{a}}{}_{a}$, so $h'^{\mathsf{a}}{}_{a}$ satisfies (1) – and we know from §4.2.2 that this is the *same* Weitzenböck derivative as that associated to $h^{\mathsf{a}}{}_{a}$ (so, in turn, the torsion tensors associated to $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ are also identical). Thus, the pair $\langle h'^{\mathsf{a}}{}_{a}, \bar{\nabla}_a \rangle$ satisfies (2)-(3).[56] For (4), note that the field equations (4.2.29) are invariant under transformations of tetrad indices, so if $\langle h^{\mathsf{a}}{}_{a}, \bar{\nabla}_a \rangle$ satisfies (4), then so too does $\langle h'^{\mathsf{a}}{}_{a}, \bar{\nabla}_a \rangle$. Thus, $\langle h'^{\mathsf{a}}{}_{a}, \bar{\nabla}_a \rangle$ also satisfies (1)-(4).

Thus, for each DPM $\mathcal{M}$ of GR, there exists a class of DPMs $\bar{\mathcal{M}}$ of TPG, differing by Lorentz transformations of the tetrad field. □

[56] The uniqueness of $\bar{\nabla}_a$ in the class of models of TPG associated to a given model of GR is in stark disanalogy with the non-uniqueness of the derivative operator in the class of models of NG related to a given model of NCT. Cf. §4.1.3.

4.2.3 Background Independence

Before assessing the background independence of TPG, first consider Knox's claim that, in this theory, "the old entities from GR appear to be waiting in the wings" [67, p. 273]. Specifically, Knox claims that in order for the tetrad field $h^{\mathsf{a}}{}_{a}$ to be defined via (4.2.2), we must specify a metric field g_{ab}; as a result, g_{ab} should be considered to have ontological priority over $h^{\mathsf{a}}{}_{a}$ in TPG. Knox claims that this is further brought out in the following points:[57] [67, p. 273]

1. The minimal coupling prescription in TPG uses the so-called *teleparallel derivative*—a recasting of the Levi-Civita derivative—rather than the Weitzenböck derivative. As a result, all non-gravitational dynamical equations in TPG take their simplest form relative to the inertial frames picked out by the Levi-Civita connection. Moreover, freely falling bodies follow geodesics of the metric field g_{ab} associated to the tetrad field via (4.2.2).
2. The tetrad field $h^{\mathsf{a}}{}_{a}$ is not uniquely determined in TPG; rather, there exists a gauge redundancy encapsulated in (4.2.8). As a result, it is tempting to think that the metric field g_{ab}—for which there is no such redundancy—should be taken as ontologically prior to the tetrad field.

In light of such results, Knox concludes that there is no underdetermination between GR and TPG; rather, the latter is simply a redescription of the former. Assuming that this is correct, then since GR is a paradigm case of a background independent theory, so too must be TPG, which is a mere reformulation of GR. One may, however, question such reasoning. For instance, while it is true that one *could* define a tetrad field from a metric field g_{ab} via (4.2.2), Knox has not demonstrated that the opposite direction is not possible – i.e. that one cannot define a metric field g_{ab} from a tetrad field $h^{\mathsf{a}}{}_{a}$. Indeed, such is possible, via (4.2.3).

[57] We have modified Knox's (2) to circumvent the issues mentioned in footnote 50.

Now consider (1) and (2). On (1), the argument for theory *identification* on this basis is too strong: a more conservative position would state that, while *read literally* GR and TPG are committed to different ontologies, the formalism of the former is more natural than that of the latter, in light of (1) and given the empirical equivalence of the two theories. On (2), it is unclear why the existence of such a gauge redundancy in the tetrad field means that the theory is better interpreted as a reformulation of GR.[58] Thus, we take it that TPG is not a straightforward reformulation of GR; understood in a flat-footedly realist sense, TPG and GR are *different theories*, with *different* ontological commitments; moreover, they should be assessed *on their own terms* with respect to qualities such as background independence.[59]

The above established, turn now in earnest to the background independence of TPG. Beginning with Def. 8, note that (4.2.29) *is* derivable from an action principle (with Lagrangian (4.2.26)); moreover, since this Lagrangian is defined in terms of $h^{\mathsf{a}}{}_{a}$ *only*, with the tetrad fields being subject to Hamilton's principle, the latter two conditions in Def. 8 appear to be satisfied, so that TPG does appear to be background independent, on Def. 8.

In fact, however, the situation here is more nuanced than this story suggests, for in presentations of the derivation of (4.2.29) from the TPG action, only the $B^{\mathsf{a}}{}_{a}$ part of $h^{\mathsf{a}}{}_{a}$ is varied (see e.g. [1, p. 177]). This might lead one to worry that $h^{\mathsf{a}}{}_{a}$ has a non-variational component, and therefore that Def. 8 is violated. Though, on the standard presentation, this is correct, recall from §3.4.2 that one *could* also vary the x^{a} part of $h^{\mathsf{a}}{}_{a}$, and obtain the same equations of motion. If one takes this approach, then Def. 8 is indeed satisfied by TPG; *unless* one also endorses the further modification to Def. 8 suggested in §3.4.2.

How does TPG fare on Belot's conception of background independence? The first thing to state here is that it is natural to take the tetrad field $h^{\mathsf{a}}{}_{a}$ in a model of the theory $\langle M, h^{\mathsf{a}}{}_{a}, \Phi\rangle$ to

[58] If $\bar{\nabla}_a$ were to possess the gauge redundancy that Knox indicates, then there would arguably be room, by analogy with Knox's reasoning in the case of NG and NCT [68], to maintain that the spacetime structure of GR is fundamental. However, we have seen in §4.2.2 (and footnote 50) therein that such redundancy is *not* present, *pace* Knox.

[59] Cf. §§4.3.1, 7.3.1.

constitute the *geometrical structure* in the theory (cf. §3.5.1). Assuming this in what follows, then, via the TPG field equations (4.2.29), one might think (by analogy with GR) that every gauge inequivalent model of the theory corresponds to a distinct geometry $h^{\mathsf{a}}{}_{a}$,[60] so that there is a one-one correspondence between the geometrical and physical degrees of freedom in the theory, and the theory is fully background independent, as per Def. 10.

There exists one technical obstacle to this appraisal of TPG as satisfying Def. 10. We saw in (4.2.8) that there exits an 'internal' Lorentz symmetry of the tetrad indices, so that if $\langle M, h^{\mathsf{a}}{}_{a}, \Phi\rangle$ is a model of TPG, then so too is $\langle M, h'^{\mathsf{a}}{}_{a}, \Phi\rangle$, where $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ are related by a Lorentz transformation, as per (4.2.8). As in §4.1.4, there exist four open interpretive options when faced with this multiplicity of models. First, one may regard $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ as equivalent geometries, and their respective models as gauge equivalent (indeed, this is plausible, since $\bar{\Gamma}^{\rho}{}_{\mu\nu}$ and $\bar{T}^{\rho}{}_{\mu\nu}$ are *invariant* under such 'internal' Lorentz transformations). Second, one may regard $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ as *inequivalent* geometries, and their respective models as not gauge equivalent. In either of these cases, TPG's satisfaction of Def. 10 is not threatened.

Alternatively, one may regard $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ as equivalent geometries, but $\langle M, h^{\mathsf{a}}{}_{a}, \Phi\rangle$ and $\langle M, h'^{\mathsf{a}}{}_{a}, \Phi\rangle$ as gauge inequivalent; as in our discussion of NG in §4.1.4, such a move does not appear defensible, for such models differ in $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ only. Finally, one may regard $h^{\mathsf{a}}{}_{a}$ and $h'^{\mathsf{a}}{}_{a}$ as inequivalent geometries, yet consider their models to be gauge equivalent. Whether such a move is ultimately plausible[61] is questionable; however, if it *is* endorsed, then Belot's Def. 10 is not satisfied.

Finally, consider whether TPG satisfies Def. 5 and Def. 6. One may begin by asking the following question: What is the status of the metric field η_{ab} in (4.2.2)? Since η_{ab} is fixed to be $\mathrm{diag}\,(-1,1,1,1)$ in all models of TPG, it is appropriate to regard this object as a *confined object*.[62] Since it is *not* the case that all confined objects should be considered absolute objects

[60] Modulo the analogue of the worries that arise for GR in this regard (cf. §3.5.4).
[61] Or indeed coherent, in light of footnote 31.
[62] Indeed, we saw in §3.4.2 that such was a viable strategy when approaching parameterised theories.

(for then even GR would manifest a large range of absolute objects – see §3.3.2), on this interpretation, in spite of the existence of this fixed structure in TPG,[63] η_{ab} does not lead to violations of background independence, on Def. 5 or Def. 6.[64]

To close, consider Pitts' point that use of the tetrad formalism in any theory gives rise to the presence of an absolute object [95, p. 365]. Given both local Lorentz freedom (in the sense of (4.2.8)) and coordinate freedom, one can bring the $\mathsf{a} = 0$ component of the tetrad field, $h^0{}_a$, into the form $(1, 0, 0, 0)$ in the neighbourhood of any $p \in M$; moreover, there cannot be any spacetime region in any model such $h^0{}_a$ vanishes. Thus, TPG (or, in the example upon which Pitts focuses [95, p. 365], GR coupled to a spinor field using an orthonormal tetrad field) gives an example of Jones-Geroch vector field: nowhere vanishing, everywhere timelike, and gauge equivalent to $(1, 0, 0, 0)$ – and so in turn, an example of an Andersonian absolute object.[65]

Faced with this example, Pitts demonstrates how, by moving to the alternative Ogievetsky-Polubarinov spinor formalism [86], 'irrelevant variables' may be excised, and this absolute object may be avoided [95, 98]. If such a move can be made in the context of TPG,[66] then this theory's having an absolute object in this regard is avoided, thereby lifting a possible roadblock to its satisfaction of Def. 5. If such a move cannot be made, however, then TPG does violate Def. 5.

4.3 Kaluza-Klein Theories

In the above, we assessed the background independence of NG (§4.1.2), NCT (§4.1.3), and TPG (§4.2.2). One further classical theory worthy of consideration is *Kaluza-Klein theory*

[63]Which, if we are to follow Smolin [123, p. 204], should ultimately be excised from our physics – see §3.3.2.

[64]Though below we see that Def. 5 is violated in TPG for different reasons.

[65]As Pitts states, "The tetrad-spinor example seems rather more serious a problem for definitions of absolute objects than the Jones-Geroch cosmological dust example was, because the spinor field is surely closer to being a fundamental field than is dust or any other perfect fluid." [95, p. 365].

[66]Without the resulting theory being distinct from TPG.

(KKT). We introduce KKT in §4.3.1, before appraising its background independence in §4.3.2.

4.3.1 Kaluza-Klein Theory

Transformations

KKT—introduced by Kaluza in 1921 [64], and developed by Klein in 1926 [66]—is an attempt to unify gravity and electromagnetism by construction of a higher-dimensional theory. KPMs of this higher dimensional theory are pairs $\langle M, \tilde{g}_{AB}\rangle$, where M is a five-dimensional differentiable manifold, and $\tilde{g}_{AB}$ is a Lorentizan metric field on M.[67] DPMs of KKT are picked out via the five-dimensional vacuum Einstein equations

$$\tilde{R}_{AB} = 0. \tag{4.3.1}$$

Here, $\tilde{R}_{AB}$ is the Ricci tensor associated to $\tilde{g}_{AB}$. Upon this metric field, we impose in every coordinate basis (following Kaluza [64]) the *cylinder condition*,[68]

$$\partial_5 \tilde{g}_{\Lambda\Sigma} = 0. \tag{4.3.2}$$

Imposition of this condition carries a number of important consequences [41, pp. 70-71].

[67] We use capital Roman indices for abstract five-dimensional indices, and capital Greek indices for five dimensional indices in a coordinate basis; to denote a set of five-dimensional coordinates, we sometimes use the notation $\tilde{x} = \left(x, x^5\right)$, where x denotes the first four coordinates in $\tilde{x}$.

[68] One understands (4.3.2) both as a restriction on the allowed coordinate systems in KKT—in analogy with imposition of the unimodular coordinate condition $\sqrt{-g} = 1$ in UGR (cf. footnote 17 of §3.3.3)—and as a restriction on $\tilde{g}_{AB}$.

First, generically we may parameterise a metric field $\tilde{g}_{AB}$ which satisfies this condition as

$$\tilde{g}_{AB} = \begin{pmatrix} g_{ab}(x) + \Phi(x) A_a(x) A_b(x) & \Phi(x) A_a(x) \\ \Phi(x) A_b(x) & \Phi(x) \end{pmatrix}; \tag{4.3.3}$$

the important point to note here is that all terms in (4.3.3) depend upon x only (not x^5). Second, if (4.3.2) is to be satisfied in every frame, then the most general allowed coordinate transformations are

$$x^\mu \to y^\mu = f(x^\mu), \tag{4.3.4}$$

$$x^5 \to y^5 = x^5 + g(x^\mu). \tag{4.3.5}$$

To see how this latter result is obtained, first expand $\partial_\Lambda \tilde{g}_{\Sigma\Theta}$ in a new coordinate basis (indices of which are denoted by primes):

$$\frac{\partial}{\partial x^\Lambda}\tilde{g}_{\Sigma\Theta} = \frac{\partial^2 y^{\Sigma'}}{\partial x^\Lambda \partial x^\Sigma}\frac{\partial y^{\Theta'}}{\partial x^\Theta}\tilde{g}_{\Sigma'\Theta'} + \frac{\partial y^{\Sigma'}}{\partial x^\Sigma}\frac{\partial^2 y^{\Theta'}}{\partial x^\Lambda x^\Theta}\tilde{g}_{\Sigma'\Theta'} + \frac{\partial y^{\Lambda'}}{\partial x^\Lambda}\frac{\partial y^{\Sigma'}}{\partial x^\Sigma}\frac{\partial y^{\Theta'}}{\partial x^\Theta}\frac{\partial}{\partial y^{\Lambda'}}\tilde{g}_{\Sigma'\Theta'}. \tag{4.3.6}$$

Hence

$$\frac{\partial}{\partial x^5}\tilde{g}_{\Sigma\Theta} = \frac{\partial^2 y^{\Sigma'}}{\partial x^5 \partial x^\Sigma}\frac{\partial y^{\Theta'}}{\partial x^\Theta}\tilde{g}_{\Sigma'\Theta'} + \frac{\partial y^{\Sigma'}}{\partial x^\Sigma}\frac{\partial^2 y^{\Theta'}}{\partial x^5 x^\Theta}\tilde{g}_{\Sigma'\Theta'} + \frac{\partial y^{\Lambda'}}{\partial x^5}\frac{\partial y^{\Sigma'}}{\partial x^\Sigma}\frac{\partial y^{\Theta'}}{\partial x^\Theta}\frac{\partial}{\partial y^{\Lambda'}}\tilde{g}_{\Sigma'\Theta'}. \tag{4.3.7}$$

By the cylinder condition (4.3.2), the left hand side of (4.3.7) is required to vanish; as a result, we also expect the terms of the right hand side of (4.3.7) to vanish term-by-term (in the general situation, no cross-term cancellation is possible). First considering the third term, we accordingly require that

$$\frac{\partial y^{\mu}}{\partial x^{5}}\frac{\partial y^{\Sigma'}}{\partial x^{\Sigma}}\frac{\partial y^{\Theta'}}{\partial x^{\Theta}}\frac{\partial}{\partial y^{\mu}}\tilde{g}_{\Sigma'\Theta'}+\frac{\partial y^{5}}{\partial x^{5}}\frac{\partial y^{\Sigma'}}{\partial x^{\Sigma}}\frac{\partial y^{\Theta'}}{\partial x^{\Theta}}\frac{\partial}{\partial y^{5}}\tilde{g}_{\Sigma'\Theta'}\stackrel{!}{=}0. \tag{4.3.8}$$

The second term here is also required to vanish by the cylinder condition (4.3.2). For the first term to vanish, we require that

$$\frac{\partial y^{\mu}}{\partial x^{5}}\stackrel{!}{=}0 \qquad \Rightarrow \qquad y^{\mu}=f\left(x^{\mu}\right), \tag{4.3.9}$$

which is (4.3.4).

Now consider the first and second terms on the right hand side of (4.3.7) – in fact, by symmetry, it suffices to consider just one of these terms.

$$\frac{\partial^{2}y^{\Theta'}}{\partial x^{5}\partial x^{\mu}}\frac{\partial y^{\Sigma'}}{\partial x^{\nu}}\tilde{g}_{\Theta'\Sigma'}=\frac{\partial^{2}y^{5}}{\partial x^{5}\partial x^{\mu}}\frac{\partial y^{\sigma}}{\partial x^{\nu}}\tilde{g}_{5\sigma}+\frac{\partial^{2}y^{5}}{\partial x^{5}\partial x^{\mu}}\frac{\partial y^{5}}{\partial x^{\nu}}\tilde{g}_{55}=0 \tag{4.3.10}$$

$$\frac{\partial^{2}y^{\Theta'}}{\partial x^{5^{2}}}\frac{\partial y^{\Sigma'}}{\partial x^{\nu}}\tilde{g}_{\Theta'\Sigma'}=\frac{\partial^{2}y^{5}}{\partial x^{5^{2}}}\frac{\partial y^{\sigma}}{\partial x^{\nu}}\tilde{g}_{5\sigma}+\frac{\partial^{2}y^{5}}{\partial x^{5^{2}}}\frac{\partial y^{5}}{\partial x^{\nu}}\tilde{g}_{55}=0 \tag{4.3.11}$$

(Setting $\Sigma=\Theta=5$ does not yield any new, independent conditions.) From (4.3.11), we extract the condition

$$\frac{\partial^2 y^5}{\partial {x^5}^2} \stackrel{!}{=} 0 \qquad \Rightarrow \qquad y^5 = \kappa\left(x^\lambda\right) x^5 + g\left(x^\lambda\right), \tag{4.3.12}$$

while from (4.3.10) we extract the condition

$$\frac{\partial^2 y^5}{\partial x^5 \partial x^\mu} = \frac{\partial}{\partial x^\mu}\left(\frac{\partial y^5}{\partial x^5}\right) = \frac{\partial \kappa\left(x^\lambda\right)}{\partial x^\mu} \stackrel{!}{=} 0, \tag{4.3.13}$$

from which we infer $\kappa\left(x^\lambda\right) = \mathrm{const}$; finally, rescaling (4.3.12) and absorbing the factor of κ^{-1} into $g\left(x^\lambda\right)$ yields (4.3.5).

Field Equations

Thus far, KKT is equivalent to five-dimensional vacuum GR, with one auxiliary condition, (4.3.2). Given the form of the metric (4.3.3) which results from this condition, one must ask what the field equations (4.3.1) of the theory become, when (4.3.3) is substituted thereinto. If one were to do so at this point, one would obtain three dynamical equations which feature non-trivial couplings between g_{ab}, A^a and Φ.[69] In KKT, however, the goal is to obtain dynamical equations which represent coupled Maxwell-Einstein gravity *only*. In order to attain this result, one imposes at this stage a second constraint: $\Phi = 1$, thus fixing this degree of freedom in the theory [90, p. 293].[70] Then, rescaling $A^a \rightarrow 4\sqrt{\pi}A^a$, the (a, b)-components of (4.3.1) become

$$G_{ab} = 8\pi T^{EM}_{ab}, \tag{4.3.14}$$

[69] See e.g. [91, p. 1].
[70] Cf. footnote 68.

with G_{ab} the 'Einstein tensor' associated to the g_{ab} components of $\tilde{g}_{AB}$, and

$$T_{ab}^{EM} = \frac{1}{4} g_{ab} F_{cd} F^{cd} - F_a{}^c F_{bc} \tag{4.3.15}$$

the stress-energy tensor of electromagnetism. The $(a,5)$-components of (4.3.1) become

$$F_{ab}{}^{;a} = 0, \tag{4.3.16}$$

with $F_{ab} = 2A_{[b;a]}$ the Faraday tensor. Thus, the field equations (4.3.1) of KKT collapse into the Einstein field equations of GR, plus the source-free Maxwell equations, coupled together.

Diffeomorphism Invariance

Consider the transformation of the $\tilde{g}_{\mu 5}$ piece of $\tilde{g}_{AB}$ (in some arbitrary coordinate basis) from one coordinate basis to another, subject to (4.3.2). We have, from (4.3.4) and (4.3.5),

$$\begin{aligned} \tilde{g}_{\mu 5} \to \tilde{g}_{\mu 5} \frac{\partial y^5}{\partial x^5} + \tilde{g}_{55} \frac{\partial y^5}{\partial x^\mu} \frac{\partial y^5}{\partial x^5} \\ = \tilde{g}_{\mu 5} + \tilde{g}_{55} \partial_\mu g\left(x^\lambda\right), \end{aligned} \tag{4.3.17}$$

so that

$$A_\mu \to A_\mu + \partial_\mu g\left(x^\lambda\right). \tag{4.3.18}$$

(4.3.18) is the transformation law for the electromagnetic vector potential under a $U(1)$ rotation. Since this result is a consequence of (4.3.5), it is sometimes written that (4.3.2) breaks the invariance group of diffeomorphisms of KKT from $\mathrm{Diff}(M)$ to a subgroup $\mathrm{Diff}(N) \times U(1)$, where N is a four-dimensional manifold; the $\mathrm{Diff}(N)$ invariance is identified with transformations (4.3.4); the $U(1)$ invariance is identified with transformations (4.3.5).

In this thesis, we avoid speaking in these terms, for three reasons. First, it is incorrect to state that (4.3.18) follows directly from (4.3.5) – rather, it follows as a result of (4.3.4) and (4.3.5) *together*. Thus, it is not clear that one should identify (4.3.5) with the $U(1)$ invariance encapsulated in (4.3.18). Second, the above reasoning leaves ill-specified at a mathematical level why (4.3.5) itself corresponds to $U(1)$ transformations. Third, the manifold associated to the Lie group $U(1)$ is S^1, the sphere. Speaking of $U(1)$ transformations at this point is apt to yield confusion with Klein's *compactification conditions*, introduced in [66], in which the x^5 coordinate is rendered periodic (sometimes, authors speak of 'compactification on S^1').[71] Note, however, that *no such compactification conditions have yet been introduced*.

For our purposes, the relevant observation is that (4.3.2) restricts the class of diffeomorphisms under which the theory is invariant to (4.3.4) and (4.3.5). Thus, treating (4.3.2) as a field equation, if $\langle M, \tilde{g}_{AB} \rangle$ is a DPM of the theory, then $\langle M, d^* \tilde{g}_{AB} \rangle$ generically will not be, for arbitrary $d \in \mathrm{Diff}(M)$. This means that KKT is *not* diffeomorphism invariant, on Def. 3.[72] Finally, note that $\tilde{g}_{55}$ is *invariant* under (4.3.4) and (4.3.5).

[71] These conditions are discussed below.

[72] One alternative is to treat (4.3.2) as restricting the *KPMs* of KKT (cf. §§3.4.2, 7.3.1). On this approach, it remains the case that KKT is not diffeomorphism invariant, for if $\langle M, \tilde{g}_{AB} \rangle$ is a DPM of the theory, then on this understanding $\langle M, d^* \tilde{g}_{AB} \rangle$ will generically not even be a KPM.

Justifying the Cylinder Condition

When Kaluza introduced the above theory in 1921 [64], (4.3.2) was not given a physical motivation. In his 1926 development of the theory, Klein attempted to rectify this situation, arguing that such an explanation can be found if the x^5 coordinate is rendered *periodic* [66].[73] In doing so, we declare that points along x^5 which differ by a quantity $2\pi R_{\text{KK}}$ (where R_{KK} is the hitherto-undetermined *Kaluza-Klein radius*) are identical: [112, p. 2]

$$x^5 \sim x^5 + 2\pi R_{\text{KK}}. \tag{4.3.19}$$

Now, for illustration, consider a massless real scalar field in five dimensions, $\phi\left(x, x^5\right)$. For such a field, we make the following Fourier expansion on the periodic x^5 dimension [88, §4.1]

$$\phi\left(x, x^5\right) = \sum_{n=-\infty}^{\infty} \phi^n\left(x\right) e^{inx^5/R_{\text{KK}}}. \tag{4.3.20}$$

Then, the associated equation of motion—here the Klein-Gordon equation—reads

$$\left(\partial^\mu \partial_\mu - \frac{\partial}{\partial x^{5^2}}\right) \phi\left(x, x^5\right) = 0, \tag{4.3.21}$$

which yields the equations of motion for the Fourier components $\phi^n\left(x\right)$

$$\left(\partial^\mu \partial_\mu + m_n^2\right) \phi^n\left(x\right) = 0, \tag{4.3.22}$$

[73] Sometimes, this condition is referred to as *compactification on* S^1; we avoid such terminology. Cf. §4.3.1.

with

$$m_n = \frac{n}{R_{\text{KK}}}. \tag{4.3.23}$$

From (4.3.22) and (4.3.23), we observe that the only massless field is $\phi^0(x)$; the other fields $\phi^n(x)$ have masses of order R_{KK}^{-1}. If R_{KK} is small, then these masses are large, so in the low-energy limit, the only non-negligible Fourier component of $\phi(x, x^5)$ is $\phi^0(x)$, which is independent of x^5. In turn, this renders the physics of the $\phi(x, x^5)$ field in the five-dimensional theory effectively independent of the x^5 coordinate. Now, what holds in the above simple case of a real scalar field also applies for other fields in KKT, including $\tilde{g}_{AB}$. Once again, due to the periodicity of x^5, we can expand the components of $\tilde{g}_{AB}$ as a Fourier series,

$$\tilde{g}_{AB}\left(x, x^5\right) = \sum_{n=-\infty}^{\infty} g^n_{AB}(x)\, e^{inx^5/R_{\text{KK}}}, \tag{4.3.24}$$

whereby analogous reasoning applies (see e.g. [88, §4.1]), and the only non-negligable component of $\tilde{g}_{AB}(x, x^5)$ is $\tilde{g}^0_{AB}(x)$, which is independent of the x^5 coordinate. In this way, rendering the x^5 coordinate periodic with small R_{KK} allows for a physical explanation of (4.3.2) – though, of course, such periodicity itself remains unexplained.

Variational Principles

With the above in mind, consider now how the equations of motion (4.3.14), (4.3.15), and (4.3.16), in turn derived from the KKT equation of motion (4.3.1), are obtainable from an action principle. We begin with the five-dimensional Einstein-Hilbert action,

$$S_{\mathrm{KK}} = \frac{1}{16\pi} \int d^5x \sqrt{-\tilde{g}} R\left(\tilde{g}\right), \tag{4.3.25}$$

where here $\tilde{g} = \det\left(\tilde{g}_{AB}\right)$, and $R\left(\tilde{g}\right) = \tilde{R}$ is the Ricci scalar associated to $\tilde{g}_{AB}$. Substituting in the specific form of the metric (4.3.3), upon which (4.3.2) has been imposed, we obtain, up to an extra integral over the x^5 coordinate, the standard action for coupled Einstein-Maxwell gravity,

$$S_{\mathrm{KK}} = \frac{1}{16\pi} \int dx^5 \int d^4x \sqrt{-g} \left(R\left(g\right) + \frac{1}{4} F_{ab} F^{ab} \right). \tag{4.3.26}$$

At this stage, however, one might think that we have a problem, for the integral over the x^5 coordinate will in general diverge. However, if we again impose that this coordinate is periodic, with (by convention) period $2\pi R_{\mathrm{KK}}$, we can overcome this problem, for in this case the integral becomes a trivial constant. In that case, performing the x^5 integral in (4.3.26), we obtain [20, p. 1093]

$$S_{\mathrm{KK}} = \frac{2\pi R_{\mathrm{KK}}}{16\pi} \int d^4x \sqrt{-g} \left(R\left(g\right) + \frac{1}{4} F_{ab} F^{ab} \right), \tag{4.3.27}$$

where R_{KK} is the radius of the periodic x^5 dimension. Hence, we see that, in addition to providing an explanation for (4.3.2), imposing periodicity of the x^5 coordinate allows one to avoid divergences in the derivation of the equations of motion (4.3.14), (4.3.15), and (4.3.16) from a five-dimensional action principle.

4.3.2 Background Independence

We are now in a position to assess the background independence of KKT. The first thing to note is that—as remarked upon in §4.3.1—KKT is *not* diffeomorphism invariant, for (4.3.2) restricts the allowed class of diffeomorphisms under which the the theory is invariant to (4.3.4) and (4.3.5). Although this is a natural conclusion, which is also endorsed by authors such as Vassallo [137, p. 10],[74] it is worth questioning at this stage whether this is *necessarily* the case.

To see what one might have in mind here, recall Dewar's distinction between *Penrose* and *Earman-Friedman* approaches to NG [31, p. 324].[75] While the Earman-Friedman approach is that presented in §4.1.3, the Penrose approach proceeds differently – in this case, a classical spacetime is defined as the direct product of a three-dimensional manifold A and a one-dimensional manifold T; on such manifolds are defined separate metric fields [31, p. 318]. Proceeding in this manner, it is natural to investigate the group $\text{Diff}\,(A \times T)$, for if this is the same as the group of diffeomorphisms of NG or NCT on the Earman-Friedman approach, then it seems that the *most general* class of possible diffeomorphisms in the Penrose approach coincides with the diffeomorphisms under which the theory is invariant.

In principle, one can make a similar move in KKT, in which the pursuant of the analogue of the Penrose approach to NG or NCT begins by defining metric fields on the product manifold $N \times S^1$, before then investigating the group $\text{Diff}\,(N \times S^1)$. If this group of diffeomorphisms coincides with the group of diffeomorphisms under which KKT is invariant (i.e. the group of smooth coordinate transformations which take DPMs of KKT to DPMs – (4.3.4) and (4.3.5)), then it would follow that whether the theory is diffeomorphism invariant depends upon the formulation of the theory at play.

[74] Although in that case on the inference that the diffeomorphism group of KKT is $\text{Diff}\,(N) \times U\,(1)$, which was criticised in §4.3.1.

[75] The former since it is Penrose's choice of presentation of NG [92, ch. 17]; the latter since it is the preferred presentation of Earman [34, ch. 2] and Friedman [46, ch. 3, §1].

Unfortunately, such an approach does *not* go through. First, for two manifolds M_1 and M_2, in general $\mathrm{Diff}\,(M_1 \times M_2) \not\cong \mathrm{Diff}\,(M_1) \times \mathrm{Diff}\,(M_2)$. Second, even supposing that the diffeomorphism group of KKT *is* $\mathrm{Diff}\,(N) \times \mathrm{Diff}\,(S^1)$ (which is *not* the case), so that $\mathrm{Diff}\,(N)$ is identified with the transformations (4.3.4), the group $\mathrm{Diff}\,(S^1)$, as discussed in §4.3.1, is nontrivial, and not straightforwardly identifiable with the transformations (4.3.5).[76] Thus, KKT is not rendered diffeomorphism invariant by moving to a Penrose-style formulation.[77]

Let us turn now to whether KKT is background independent. Since KKT does not contain any fixed fields, Def. 6 is satisfied.[78] By contrast, Def. 5 is more complex. On the one hand, the component $\tilde{g}_{55}$ is invariant under (4.3.4) and (4.3.5). However, at this point this object is still formally *unfixed* – meaning that the space of DPMs of KKT is partitioned into domains, points of which are related by allowed diffeomorphisms, and in each of which $\tilde{g}_{55}$ takes a different value. This situation thus resembles Torretti's counterexample to the equation of the absence of absolute objects and background independence (§3.3.3), and therefore—one might argue—does not present a violation of Def. 5. In fact though, recall that, as Pitts points out, such a case *does* involve an absolute object – the square root of the metric determinant [95, pp. 17-18]. In addition, in the final formulation of KKT, $\tilde{g}_{55}$ is fixed via $\Phi = 1$. This extra condition restricts the DPMs of KKT to only one of the above-mentioned equivalence classes – in which case there again arises an absolute object in this example.

Consider next Def. 8. Recalling the KKT action (4.3.25), it is clear that all dependent variables therein (namely the one field, $\tilde{g}_{AB}$) are subject to Hamilton's principle, and represent physical fields. Thus, it appears that KKT qualifies as background independent, on Def. 8. One subtlety in this regard, however, is the following. Above, we suggested that the conditions (4.3.2) and $\Phi = 1$ be imposed as *field equations* – i.e. at the level of DPMs. But if this is the case, we should be able to *derive* these equations from an action principle – and it is not

[76] See e.g. [65, ch. 2].

[77] The analogue of these results also applies to the Penrose formulation of NG.

[78] Absent any reformulation of KKT containing fixed fields.

obvious that this can be achieved. In fact, only in the case in which these conditions are imposed at the level of *KPMs* does KKT satisfy Def. 8 uncontroversially (cf. §§3.4.2, 7.3.1).

Finally, turn to Belot's account of background independence. Here, one must take $\tilde{g}_{AB}$ to represent the geometrical structure (see §3.5.1) in the theory; then, any such metric fields related by the allowed transformations (4.3.4) and (4.3.5) can be considered *equivalent geometries* (as with GR). Moreover, as in the case of GR, models which differ by metric fields related by such transformations can be considered physically equivalent. In this case, KKT satisfies Def. 10, of *full* background independence.[79]

The combination of our results that KKT is *not* diffeomorphism invariant, yet on several approaches *is* background independent, is important, for it further refutes the idea that diffeomorphism invariance is intimately tied to background independence. Specifically: just as Pooley demonstrates via **SR2** that one can have a diffeomorphism invariant theory that is not background independent (see §3.2.4), in KKT we appear to have a theory which is *not* fully diffeomorphism invariant, yet which is nevertheless background independent.This result should be taken as an additional warning against the identification of these concepts.

Let us close with the following comment. As with Knox on TPG and GR (cf. §4.2.3), one might argue—based upon the fact that one can obtain the Einstein-Maxwell action (4.3.27) simply by integrating out the (periodic) x^5 coordinate in the Kaluza-Klein action (4.3.25)—that KKT is simply a reformulation of GR, rather than a distinct theory. If so, then, again, KKT should inherit the background independence of GR. As before, however, we wish to resist this reasoning: KKT is *prima facie* ontologically distinct from four-dimensional Einstein-Maxwell theory, and should be evaluated *on its own terms*.[80]

[79] One might worry that this verdict is too quick. Specifically, what of diffeomorphisms which do *not* give rise to coordinate transformations of the form (4.3.4), (4.3.5)? Since such diffeomorphisms take us *out* of the class of DPMs of the theory—and recall from §3.5.4 that background independence on Belot's account should be assessed with respect to the DPMs of the theory in question—these transformations pose no obstacle to the appraisal of full background independence in this case.

[80] In this regard, we follow Vassallo [137, pp. 10-11].

4.4 Close

In this chapter, we have assessed the background independence of NG, NCT, TPG, and KKT; our results are summarised in Table 4.1. Such work is valuable for four reasons. First, it provides a means of testing the definitions presented in §3, revealing potential shortcomings which may otherwise have gone unnoticed. For example, we saw in §4.1.4 that Def. 10 delivers the incorrect verdict on NCT, due to issues regarding the counting of *geometrical structures*.

Second, carrying out this work allows us to move beyond intuitions in appraising whether the theories under consideration are background independent. For example, one might think that NCT is background independent, since it involves a dynamical spacetime as with GR. In §4.1.4, however, we found a majority verdict that NCT is *not* background independent, in spite of intuitions. Indeed, in cases such as KKT, it is not even clear that we *have* strong intuitions regarding background independence; the majority verdict of background independence presented in §4.3.2 is therefore useful when making judgements on this theory.

Third, this work develops our understanding of the connections between background independence and other concepts. For example, in KKT we find a theory which is not diffeomorphism invariant, yet which nevertheless arguably *is* background independent. And fourth: The analysis undertaken above leads to an enhanced understanding of the theories in question. For example, we have discussed various forms of indeterminism in NG and NCT (§4.1.4); different classes of gauge transformation in TPG (§4.2.3); and issues regarding the diffeomorphism group associated to KKT (§4.3.2).

	Def. 5	Def. 6	Def. 8	Def. 10
NG	✗	✗	✗	✗*
NCT	✗	✗	✗	✓*
TPG	✗	✓	✓*	✓*
KKT	✗	✓	✓*	✓

Table 4.1: The appraisals delivered by Def. 5, Def. 6, Def. 8, and Def. 10, on the background independence of NG, NCT, TPG, and KKT. ✓ indicates that the theory is judged to be background independent; ✗ indicates that the theory is judged to be background dependent; and * indicates that there exist subtleties regarding the verdict.

Chapter 5

Quantum Background Independence

In §2.1, we presented a picture of classical theories in terms of KPMs, DPMs, and BPMs. Modifications to this framework are necessary in the case of quantum theories. In §5.1, we address these issues, before considering in §5.2 whether the definitions of background independence presented in §3 remain applicable to quantum theories.

5.1 Models, Reprise

The KPMs of a quantum field theory (QFT) are tuples $\langle M, F_i, \ldots, F_m, \hat{O}_1, \ldots, \hat{O}_n \rangle$, where the F_i are fixed fields,[1] and the $\hat{O}_i$ are *operator-valued fields*.[2] Then, in such a QFT, the relevant *dynamical* question is not 'what dynamical equations do the $\hat{O}_i$ obey?', but 'what are

[1] In the sense of **SR1** – see §3.2.2. It in necessary to include the F_i, for in e.g. standard Lorentz-covariant QFT, quantum fields are defined on a *fixed* background Minkowski spacetime.

[2] The introduction of F_i and $\hat{O}_i$ raises the possibility that one might also include in the definition of the KPMs of a QFT *dynamical* classical fields. Such a possibility is of relevance to §6 (and cf. [106]), though for simplicity is here set aside.

the correlation functions of the $\hat{O}_i$?'[3] Recall that such correlations functions

$$\langle O_{\alpha_1}(x_1)\dots O_{\alpha_m}(x_m)\rangle \equiv \langle\Omega|\, T O_{\alpha_1}(x_1)\dots O_{\alpha_m}(x_m)\,|\Omega\rangle\,, \tag{5.1.1}$$

where each $O_{\alpha_i} \in \{O_1 \dots O_n\}$, $i = 1\dots m$, give the probability amplitude of finding an excitation of field O_{α_i} at spacetime location x_i ($i = 1\dots m$), for each of the O_{α_i}.[4] Hence, for example, the two-point correlation function $\langle\varphi(x_1)\varphi(x_2)\rangle$ of some real scalar field φ can (loosely) be interpreted as the probability amplitude for 'propagation' of an excitation of the φ field (i.e., a φ-'particle') from the spacetime point with coordinates x_1, to the spacetime point with coordinates x_2.

With the above picture of correlation functions in mind, now recall that in a QFT, such correlation functions can be computed via the *path integral*

$$\int DO_1\dots DO_n \exp\left(i\int_{-\tau}^{\tau} d^4x\mathcal{L}\right) \tag{5.1.2}$$

—i.e., the functional integral of the exponential of a classical action (itself a spacetime integral over a Lagrangian $\mathcal{L}$)—using (see e.g. [94, ch. 9])

$$\langle O_{\alpha_1}(x_1)\dots O_{\alpha_m}(x_m)\rangle = \lim_{\tau\to\infty(1-i\epsilon)} \frac{\int DO_1\dots DO_n O_{\alpha_1}(x_1)\dots O_{\alpha_m}(x_m)\exp\left(i\int_{-\tau}^{\tau} d^4x\mathcal{L}\right)}{\int DO_1\dots DO_n\exp\left(i\int_{-\tau}^{\tau} d^4x\mathcal{L}\right)}. \tag{5.1.3}$$

[3] In what follows, hats on operators shall be dropped for convenience.

[4] $|\Omega\rangle$ denotes the ground state of the (interacting) QFT in question; T is a *time-ordering operator*. See e.g. [94, ch. 4]. Note that one might also choose to include the quantum state in the definition of the KPMs of a QFT; whether one does so will not prove significant for our purposes.

Now ask: How should DPMs of a QFT be picked out? The obvious answer is that the DPMs of a QFT are those KPMs for which correlation functions of the O_i are specified by a certain path integral. Thus, in short, DPMs of QFTs are picked out by path integrals[5] – this is the analogue of picking out DPMs via field equations, as in §2.1.[6]

Finally, what of the BPMs of a classical versus a quantum theory? In this case, one can impose certain boundary conditions on the O_i of a QFT in exactly the same manner as was discussed in §2.1. Thus, no modification to the formalism of that section is necessary in this context – beyond the fact that, as mentioned, the O_i are now operator-valued fields.

5.2 Background Independence

The definitions of background independence presented in §3 were all constructed in the context of *classical* field theories. It is thus an open question whether they remain applicable to QFTs. Thankfully, however, this is the case. To illustrate, note that the definition of a fixed field (i.e. a field fixed in all KPMs of a theory – see §3.2), and of an absolute object (i.e. an object invariant across all DPMs of a theory – see §3.3), are equally applicable to quantum theories as they are to classical theories. Thus, Def. 5 and Def. 6 can also be applied to assess the background independence of quantum theories. Indeed, it is also the case that no significant modifications need be made to Belot's account of background independence, beyond a reconstrual of the models of a theory *à la* §5.1.

[5]To specify a path integral, we have seen that one must specify a classical action, and so in turn a Lagrangian. However, as in classical theories, there exist some QFTs which lack a Lagrangian formulation (for example $\mathcal{N} = (0, 2)$ QFTs in six dimensions – see e.g. [24,144]). Nevertheless, the framework presented in this section is capable of accommodating such theories – we simply specify that DPMs are picked out by correlation functions *directly*, without recourse to such path integrals. Indeed, to say that path integrals pick out DPMs in QFTs is, ultimately, a convenient *shorthand* for stating that DPMs are picked out by certain correlation functions.

[6]In the language of QFT, to pick out DPMs via field equations is to pick out those models which capture only the *on-shell*, or *tree-level*, contributions to the dynamics. In a QFT, we also need to capture loop-order contributions; picking out DPMs via a path integral allows us to achieve this goal.

The situation regarding Def. 8 is a little more complex. Suppose that the theory in question has an associated action. In a quantum theory, such an action can be varied to obtain the on-shell equations of motion for the objects O_i of the theory. One might worry, however, that since it is the *path integral* (rather than the action) which is fundamental in a QFT, such variation of the action is irrelevant to any appraisals of the background independence of this theory.

Now, while it is true that such equations of motion do not give the *full* dynamics of the theory (unlike the classical case)—which is given by the path integral—what *is* true is that generic correlation functions in a QFT can be evaluated perturbatively,[7] with each term in such a perturbation series being constructed from *propagators*, themselves constructed from the equations of motions for the O_i of the theory which are obtained form such action principles. Thus, if a field in a QFT is subject to Hamilton's principle, this is good reason to think it dynamical in the quantum theory; if a field in a QFT is *not* subject to Hamilton's principle, then it does not appear to be dynamical in the quantum theory. Thus, the intuitions behind Def. 8 of background independence still hold in quantum theories – suggesting that it *is* still appropriate to make use of this criterion in such contexts.

It is worth mentioning one further modification to Def. 8 (also applicable to Def. 5 and Def. 6) which must be made when assessing the background independence of quantum theories (brought to light by de Haro *et al.* [25, §6.2][8]). Since the path integral of a quantum theory decomposes into the exponential of a classical action, *and* a path integral measure, it is important that the latter object *also* contain no non-dynamical—or non-variational, etc.—fields.

[7]In terms of sums over Feynman diagrams – see e.g. [94, ch. 4].

[8]See §7.2.1.

Chapter 6

Perturbative String Theory

In this chapter, we assess whether perturbative string theory qualifies as background independent. To this end, in §6.1 we provide an overview of this quantum gravity theory. Then, in §6.2, we consider the nature of spacetime in the theory, before turning to our main task in §6.3.

6.1 Background

In string theory, a string can be regarded as a special case of a p-brane, which is an object with p spatial dimensions and tension $T_p = 1/\left(2\pi\alpha'\right)$, where α' is the *Regge slope parameter*. The classical motion of a p-brane extremises the $(p+1)$-dimensional volume V that it sweeps out in spacetime.[1] Thus there is a p-brane action that is 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

[1] Sometimes, spacetime in string theory is referred to as *target space*. Recently, in light of so-called *dualities* in string theory (see e.g. [108]), Huggett has argued that phenomenal spacetime is *not* equivalent to target space [59]. We set such issues aside in this thesis.

called the *Nambu-Goto action* [17, p. 10]

$$S_{\rm NG} = -\frac{1}{2\pi\alpha'}\int_\Sigma d^2\sigma\left[-\det{}_{\mathfrak{ab}}\left(\frac{\partial X^a}{\partial\sigma^{\mathfrak{a}}}\frac{\partial X^b}{\partial\sigma^{\mathfrak{b}}}\eta_{ab}\right)\right]^{1/2}, \tag{6.1.1}$$

where Σ denotes the string worldsheet; the functions $X^a(\sigma,\tau)$ describe the spacetime 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. Note in (6.1.1) that the spacetime metric field η_{ab} is a *fixed field* in the sense of §3.2.2. Classically, the Nambu-Goto action is equivalent to the string sigma-model action (also known as the *Polyakov action*) [17, p. 13]

$$S_{\rm P(\eta)} = -\frac{1}{4\pi\alpha'}\int_\Sigma d^2\sigma\sqrt{-h}h^{\mathfrak{ab}}\eta_{ab}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b, \tag{6.1.2}$$

where $h_{\mathfrak{ab}}(\tau,\sigma)$ is an auxiliary[2] 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 (6.1.2) and recover (6.1.1).[3] 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.[4] Doing this, one finds that there is a conformal anomaly[5] unless the spacetime

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

[3]Varying (6.1.2) 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 (6.1.1).

[4]I.e. path integral quantisation *à la* Faddeev-Popov – see e.g. [17, §3.4].

[5]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 spacetime dimension $D = 26$ for the bosonic string, or $D = 10$ for the fermionic string. For a philosophically-oriented introduction to anomalies, see [60, §4].

dimension is $D = 26$.[6]

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,[7] 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 spacetime indices, and transforms under $SO(D-2)$ as a massless,[8] 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* [12, p. 53].

In (6.1.2), we 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 [17, p. 429][9]

$$S_{\mathrm{P}(\eta+f,B,\Phi)} = -\frac{1}{4\pi\alpha'}\int_{\Sigma} d^2\sigma\sqrt{-h}\Big(h^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b\left(\eta_{ab}+f_{ab}(X)\right) + \epsilon^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b B_{ab}(X) + \alpha'\Phi R(h)\Big), \quad (6.1.3)$$

where our 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. It is upon this

[6]As mentioned in footnote 5, 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 spacetime as in the *GS sector*) gives the critical dimension $D = 10$ [12, p. 7]. In this chapter we focus solely on bosonic string theory.

[7]See e.g. [12, p. 53].

[8]For how the masses of such states are determined, see e.g. [12, p. 53].

[9]In some textbooks (e.g. [12, p. 81]), the η_{ab} term in (6.1.3) is dropped; without a field redefinition of the form $f_{ab} \to \eta_{ab} + f_{ab}$, this is a mistake, for the background η_{ab} term is still required at this stage. Other textbooks (e.g. [17, p. 428], [100, p. 108]) do not commit this error.

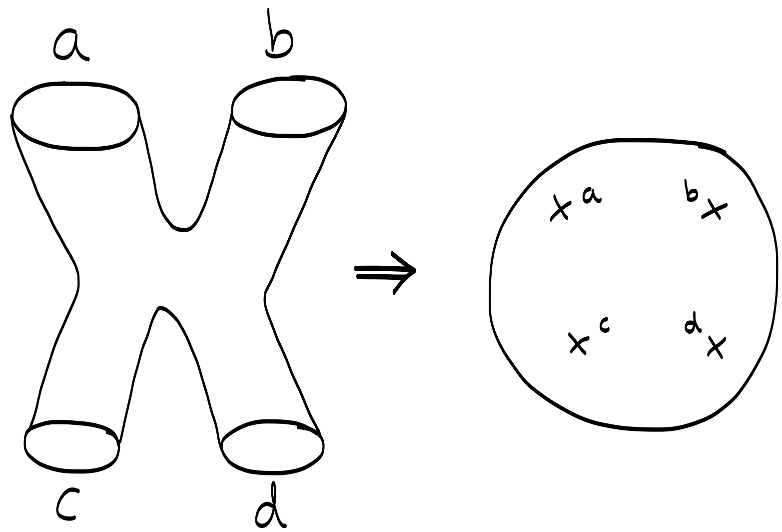

Figure 6.1: The mapping of the four-point diagram for closed strings (left) to a compactified worldsheet (topologically a sphere) in which each of the four original outgoing world tubes is mapped to a marked point (right), via the conformal invariance of the worldsheet metric.

move that we focus in §6.2.

Finally, we must discuss string interactions. In string theory, the worldsheet metric $h_{\mathfrak{ab}}$ is invariant under local changes of scale

$$h_{\mathfrak{ab}}\left(\tau,\sigma\right)\rightarrow\Omega^{2}\left(\tau,\sigma\right)h_{\mathfrak{ab}}\left(\tau,\sigma\right). \tag{6.1.4}$$

This invariance allows us to make conformal rescalings of the worldsheet metric; this is useful when evaluating string scattering amplitudes. For example, consider the four-point diagram for closed strings, shown on the left of Figure 6.1.[10] By making a suitable choice of Ω, this diagram can be converted into the second diagram, in which the worldsheet is compact, the holes in original the string worldsheet corresponding to external states are closed, and the external string states now appear at marked points.[11] After rescaling, at each marked point—

[10] Since we are now dealing with *strings* rather than point particles, we have *tubes* or *sheets* (depending on the string topology: open or closed, respectively) rather than *lines* in the analogues of Feyman diagrams.

[11] For an extended discussion of this point, see [51, pp. 32ff.].

i.e. at each string interaction vertex—there must appear a suitable local operator with quantum numbers of the string state that was mapped to that point [51, p. 34]. This is a *vertex operator*, and generically takes the form

$$V_\Theta(k) = \int d^2\sigma \sqrt{-h} W_\Theta(\tau, \sigma) e^{ik\cdot X} \tag{6.1.5}$$

for emission or absorption of string state of type Θ and momentum k^a; W_Θ is a polynomial in X^a and its derivatives [51, pp. 34ff.].[12] In particular, in the case where Θ is the graviton, we must pick the associated W_{grav} to saturate the two symmetric spacetime indices; the minimal such operator is $W_{\text{grav}} = h^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b\zeta_{ab}$, for some ζ_{ab}. Thus, we expect the vertex operator for the graviton $V_{\text{grav}}(k)$ to read

$$V_{\text{grav}}(k) = \int d^2\sigma \sqrt{-h} h^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b\zeta_{ab}e^{ik\cdot X}. \tag{6.1.6}$$

This is understood as the vertex operator for emission or absorption of a graviton of wave-function $\zeta_{ab}e^{ik\cdot X}$, and is discussed further in §6.2.

[12]The reason for the factor $e^{ik\cdot X}$ is the following. As stated at [51, p. 35], though the operators W_Θ transform correctly under Lorentz transformations, we must also take account of spacetime translations. Under the global symmetry $X^a \to X^a + a^a$ in which the position of each string is shifted by a constant a^a, the wavefunction of an external state of momentum k^a is multiplied by $e^{ik\cdot a}$. The simplest quantum field operator that transforms in this way under translations is $e^{ik\cdot X}$, so we postulate that this factor is present for emission or absorption of a string of momentum k^a.

6.2 Fields

6.2.1 Background Fields and Coherent String States

One might reasonably ask: How does one bridge the conceptual gap between f_{ab}, B_{ab} and Φ being excited states of a closed string, to these objects being *background fields*, in an action such as (6.1.3)? Heuristic attempts to answer this question have been put forward by e.g. Polchinski [100, p. 108] and Green, Schwarz and Witten [51, pp. 165-166]; let us consider the approach of the latter. Focussing on the case of augmenting (6.1.2) with an $f_{ab}\left(X\right)$ piece (i.e. ignoring for the time being contributions from B_{ab} and Φ),[13] the authors of [51] motivate the transition $S_{\mathrm{P}(\eta)} \to S_{\mathrm{P}(\eta+f)}$ by first supposing that the spacetime metric is

$$g_{ab}\left(X\right) = \eta_{ab} + f_{ab}\left(X\right), \tag{6.2.1}$$

where $f_{ab}\left(X\right)$—treated as a perturbation—represents deviation from Minkowski spacetime. The string worldsheet path integral associated to $S_{\mathrm{P}(\eta)}$—i.e. for a string in a Minkowski background—is

$$Z_{\eta} = \int DX^{a} Dh_{\mathfrak{ab}} e^{-S_{\mathrm{P}(\eta)}}, \tag{6.2.2}$$

while that associated to $S_{\mathrm{P}(g)}$ is

[13] As Huggett states [60, p. 7], dependency upon the X^{a} is now included so that the operator transforms correctly under translations. Cf. footnote 12.

$$
\begin{aligned}
Z_g &= \int DX^a Dh_{\mathfrak{ab}} e^{-S_{\mathrm{P}(g)}} \\
&= \int DX^a Dh_{\mathfrak{ab}} e^{-S_{\mathrm{P}(\eta)}} \left(1 + \frac{1}{2\pi} \int d^2\sigma \sqrt{h} h^{\mathfrak{ab}} \partial_{\mathfrak{a}} X^a \partial_{\mathfrak{b}} X^b f_{ab}(X) \right. \\
&\quad \left. + \frac{1}{2} \left[\frac{1}{2\pi} \int d^2\sigma \sqrt{h} h^{\mathfrak{ab}} \partial_{\mathfrak{a}} X^a \partial_{\mathfrak{b}} X^b f_{ab}(X) \right]^2 + \dots \right).
\end{aligned} \tag{6.2.3}
$$

The crucial step in the argument is then the following: the factor

$$
V_f = \frac{1}{2\pi} \int d^2\sigma \sqrt{h} h^{\mathfrak{ab}} \partial_{\mathfrak{a}} X^a \partial_{\mathfrak{b}} X^b f_{ab}(X) \tag{6.2.4}
$$

appearing in the above is the vertex operator for the emission or absorption of a graviton with wavefunction $f_{ab}(X)$![14] An insertion of V_f into the path integral (6.2.2) accommodates the interaction of strings with an external graviton of wavefunction $f_{ab}(X)$; an insertion of e^{V_f} into (6.2.2) describes the interaction of such a string with a coherent state of gravitons, and "this corresponds precisely to string propagation in the metric $g_{ab}(X) = \eta_{ab} + f_{ab}(X)$" [51, p. 166][15,16] As Huggett states, "A coherent state is very 'unquantum', in the sense that adding and subtracting field quanta does not affect it" [60, p. 7];[17] indeed, it is in this manner that *classical* background fields arise in perturbative string theory.[18]

[14]As stated at [51, pp. 165-166], in (6.1.6) we considered only gravitons the wave functions of which are plane waves, of the form $\zeta_{ab} e^{ik\cdot X}$. However, "there is no reason not to consider a wavefunction instead that is a general superposition of plane waves" – such as is the case for $f_{ab}(X)$.

[15]Notation in this quotation has been modified for consistency with this thesis; there is no change in content.

[16]There exist questions in this vicinity regarding whether the full space of solutions of GR can be built up from a linearised theory in the manner discussed here – see, for example, the classic work of Feynman [44], and ensuing more modern discussions such as [89, 99]. In [123, p. 24], Smolin argues that it is *not* the case that one can construct the full space of solutions of GR in perturbative string theory. Even if this were so, however, it is unclear (absent some further argument) why this should be construed as a problematic. Indeed, in a parallel manner, Barbour has argued that it is an *advantage* of his alternative programme to GR, *shape dynamics*, that it has a more restricted space of solutions than GR, for this renders the theory "more predictive" (see e.g. [8]). Though interesting, these issues shall here be set aside.

[17]See [60, §2] for an extended discussion of this point.

[18]Excluding the η_{ab} field at this point – see the discussion below.

On the basis of these results, Huggett and Vistarini conclude: [61, p. 1169][19]

> 'Background' fields do not represent new degrees of freedom in addition to those of string theory: they are not distinct primitive entities. Instead they represent the behaviour of coherent states of string excitations: the quantum states, that is, which describe classical field behaviour.

One must question whether this is strictly correct, for the metric field $g_{ab}(X)$, constructed as per (6.2.1), at this stage remains a *composite object*, consisting both of graviton excitations $f_{ab}(X)$, and an *ontologically primitive* background field η_{ab}. In order to vindicate claims such as the above, one must also account for the η_{ab} field. In fact, however, to do so is not straightforward. Naïvely, one might expand the $e^{-S_{\mathrm{P}(\eta)}}$ piece in (6.2.3),[20] and attempt to consider this also as a coherent state of gravitons. However, this move will not work, for the η_{ab} field, unlike the $f_{ab}(X)$, is *not* trace-free.[21]

As a possible way out here, one might be tempted to argue as follows. When exploring the closed string spectrum, the first excited state is decomposed into irreducible representations of the transverse rotation group $SO(D-2)$, which are then understood to correspond to the f_{ab}, B_{ab} and Φ fields.[22] However, there is no *a priori* reason why this decomposition must be complete – suppose instead that the symmetric, trace-free piece f_{ab} and trace piece Φ are *not* so split, so that one considers one first excited state of the closed string to transform under $SO(D-2)$ as a massless, spin-two particle w_{ab}, *with trace*. Then, the vertex operator for this

[19] See also [60, §§2-3], and works by string theorists, such as [57, 121, 145].

[20] With respect to the $e^{-S_{\mathrm{P}(f)}}$ piece, i.e. run the same reasoning as above for this other term in the path integral.

[21] *Pace* Huggett and Vistarini [61, p. 1167], this is also true of a "general Lorentzian metric". Note also that speaking of such 'general Lorentzian metics' consisting in coherent string states overlooks possible complications arising out of the discussion in footnote 16.

[22] For details, see e.g. [17, pp. 49-50].

object will read

$$V_w = \frac{1}{2\pi}\int d^2\sigma\sqrt{h}h^{\mathfrak{ab}}\partial_{\mathfrak{a}}X^a\partial_{\mathfrak{b}}X^b w_{ab}(X)\,, \tag{6.2.5}$$

and therefrom a background Minkowski metric field η_{ab} *can*—it is argued—be constructed, via the above argument. Again, however, this approach faces problems, for even to *identify* the first excited states of a closed string in this manner[23] *presupposes* the existence of a Minkowski spacetime metric field η_{ab} (see e.g. [17, ch. 2]). Thus, it does not appear that one can run an argument of this type without prior commitment to the existence of such a field. Consequently, we conclude at this stage that more needs to be done to bring Huggett and Vistarini's above claim to fruition – and that any story that one *does* tell here will be non-trivial in nature.

6.2.2 Primitive Worldsheet Ontology

In spite of the above scepticism, it is interesting to consider what would follow *if* a string-theoretic ontological reduction *à la* Huggett and Vistarini could successfully be carried out. Supposing this to be so, it is the worldsheet metric $h_{\mathfrak{ab}}$ that is fundamental, *not* the spacetime metric g_{ab}. There exist two obvious attractions to this mode of thinking:

1. On this account, *strings come first*; familiar ontology is built up from string states with no prior assumptions about spacetime of a 'non-stringy' nature.

2. The field $h_{\mathfrak{ab}}$ is invariant under (6.1.4), so a string theory can be considered a two dimensional conformal field theory on the worldsheet. Thus, *only conformal worldsheet structure need be postulated*.[24] This means that less metric structure is required in string theory than might be anticipated – see §6.3.

[23] And indeed in the standard manner, in terms of f_{ab}, B_{ab} and Φ fields.

[24] In fact, recent work in the foundations of GR has demonstrated an analogous result in that case [21].

6.2.3 Field Equations

Setting aside the above issues of ontological priority, we should now return to the question of whether perturbative string theory qualifies as background independent. Here, Huggett and Vistarini write the following: [61, p. 1169]

> [I]n a central sense, string theory is background independent: the metric arises from string interactions, rather than being stipulated a priori. Just like general relativity, many solutions are possible, but matter and gravity have to satisfy a mutual dynamics – except in string theory, there is no fundamental distinction between the two, a significant ontological unification.

To see what is meant here, consider again a string propagating in background fields, as per e.g. (6.1.3). As a worldsheet QFT, string theory is conformally invariant. 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.[25] Thus, the 'background fields' cannot be fixed after all—they *must* be dynamically coupled via Einstein-type field equations.

6.3 Background Independence

In light of this result, let us return to our four definitions of background independence—Def. 5, Def. 6, Def. 8 and Def. 10—in order to assess whether perturbative string theory qualifies as background independent, as Huggett and Vistarini claim. For completeness, we carry out our appraisals both at the level of spacetime fields, and at the level of worldsheet fields.

[25] For original sources on this result, see e.g. [22]; for textbook discussion, see [51, pp. 167ff.]; for a presentation in the philosophical literature, see [59, §3].

Beginning with the former, the fields obey, as mentioned in §6.2.3, the dynamical equations of GR, plus corrections. For this reason, Def. 6 and Def. 10 are straightforwardly satisfied: as with GR, there are no fixed fields in the theory governing the dynamics of these fields; moreover, there appears to be a one-one match between physical and geometrical degrees of freedom.[26] That being said, it is of course also the case that Def. 5 is *not* satisfied by these background fields, for, as with GR (cf. §3.3.3), the square root of the metric determinant will qualify as an absolute object in these theories. The situation regarding Def. 8 is a little more subtle, for one must be clear on the relevant action in this case. This is *not* the worldsheet action (6.1.3); rather, the fact that spacetime fields obey the Einstein equations plus corrections means that there exists an action, variation of which yields these equations: simply the Einstein-Hilbert action, plus corrections; *this* is the relevant action in this context. In such an action, all dependent variables are subject to Hamilton's principle, and there exist no unphysical fields. Thus, again, it *does* appear that Def. 8 is satisfied.

Turn now to the worldsheet action (6.1.3).[27] Here, again, there exist no fixed fields, absolute objects, or non-variational fields, for both $h_{\mathfrak{ab}}$ and the X^a are varied to yield associated equations of motion (see e.g. [12, p. 27]). Though the fields g_{ab}, B_{ab} and Φ might *appear* to be fixed, from the worldsheet perspective these are viewed as functions of the X^a;[28] thus, they do *not* compromise background independence on Def. 5, Def. 6,[29] or Def. 8.[30] While one might worry that the X^a fields violate the prohibition on *unphysical* fields in Def. 8, this is not so – for these functions encode precise physical information on the spacetime location of the string. In a parallel manner, while one might worry that the field $h_{\mathfrak{ab}}$ is unphysical, whether *this* is so will depend upon whether one considers $h_{\mathfrak{ab}}$ a fundamental field (§6.2.2), or

[26]Modulo the potential problem cases for GR discussed in §3.5.4, which carry across to this context.

[27]Strictly, this is just one example of a string worldsheet action featuring background fields – however, the analysis presented here carries across more generally.

[28]Which, as we saw in §6.2.3, obey precisely specifiable relations.

[29]Again remembering that we have not demonstrated that there does not exist any reformulation of the theory which *does* involve fixed fields – which would pose a problem for Def. 6.

[30]Though initially we stated that η_{ab} in (6.1.1) and (6.1.2) was considered to be a fixed field, in light of the discussion in §6.2.3, it is better to view this field as one *solution* of the more general case. Thus, this field does not compromise background independence.

an auxiliary construction (§6.1) – on the former view, the field is physical; on the latter, it is not. Whether one considers Def. 8 to be satisfied hinges upon this point.[31]

Finally, turn to Belot's account. From the worldsheet point of view, it is natural to take the field $h_{\mathfrak{ab}}$ to constitute *geometrical structure* in the theory. But now recall again that it is possible to eliminate $h_{\mathfrak{ab}}$ from (6.1.2), to recover (6.1.1). In this case, one can envisage models of perturbative string theory which differ in the X^a, so that the spacetime embedding of the string is different, but with the *same* $h_{\mathfrak{ab}}$ field on the worldsheet – and the most natural reading of this situation is as a violation of Def. 10 – *unless* one can also argue that models which differ solely in the spacetime location of the string are gauge-equivalent.[32]

[31]Since (6.1.2) is equivalent to (6.1.1), however, it may be possible to argue that even if $h_{\mathfrak{ab}}$ *is* unphysical, its presence does not compromise background independence on Def. 8 – for such appraisals should be made with respect to the latter action.

[32]One might question this verdict on the grounds that the relevant issue from the worldsheet point of view appears to be whether each non-equivalent cofiguration of geometrical structure *on the worldsheet* is considered physically non-equivalent (recall Def. 10) – with the spacetime location of the worldsheet being *irrelevant* to this appraisal. If this move is defensible, then perhaps there *is* room to argue that Def. 10 is not violated, in this context.

Chapter 7

Holography

Recently, it has been claimed that background independent gravity theories can be constructed via so-called *holographic dualities*. The idea is that one begins with a Yang-Mills QFT, then *translates* into the language of a gravity theory via the duality. In this chapter, we assess whether the gravity theories so constructed are indeed background independent. The structure is as follows. In §7.1, we introduce holographic dualities. In §7.2, we discuss two notions of background independence due to de Haro *et al.* [25, 27], which have been proposed in this context. In §7.3, we assess the background independence of such gravity theories: we find that one's assessment in this regard hinges upon the definition of background independence at play – though there exists a majority verdict that such theories *are* background independent.

7.1 Holographic Dualities

The idea of holographic duality was first proposed by 't Hooft in 1993 [56]. A holographic duality exists when a theory defined on a D-dimensional spacetime (the 'bulk') is theoretically equivalent[1] to a theory defined on a $(D-1)$-dimensional spacetime that forms the boundary of the bulk.[2] Often, the bulk theory is a gravity theory, whereas the boundary theory is a (quantum) gauge theory.[3] In that case, a holographic duality is referred to as a *gauge/gravity duality*. Such dualities are often not only mathematically striking, but also of practical value to physicists;[4] as a result, they are studied widely in contemporary mathematical physics.

Throughout this chapter, we use as a running example the best known case of gauge/gravity duality: the *AdS/CFT correspondence* (first proposed by Maldacena in 1997 [79]), in which the bulk theory is set in anti-de Sitter (AdS) spacetime,[5] and the boundary theory is a conformal field theory (CFT). One concrete example of this correspondence is that between type IIB string theory[6] on the ten-dimensional product space $AdS_5 \times S^5$—this being a putative quantum theory of gravity—and $\mathcal{N} = 4$ supersymmetric Yang-Mills theory[7] on the four-dimensional boundary of the AdS_5 space—this being a supersymmetric CFT.[8]

Let us give further detail on bulk theories in the AdS/CFT correspondence. Importantly, the low-energy approximation to the bulk equations of motion is given by Einstein's equations with negative cosmological constant, $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$ $(\Lambda < 0)$ [27, p. 14]. For any space which satisfies these equations, and given a conformal metric at infinity,[9] the associated line

[1]See footnote 3 of §4.1.

[2]That the physics of the bulk is encoded in the boundary is reminiscent of holography, hence the appellation.

[3]Care regarding the meaning of 'gauge theory' is needed – typically, in the context of gauge/gravity duality, what is meant are Yang-Mills theories, thereby excluding other candidate 'gauge theories' such as GR. (See [139] for a discussion regarding the parallels between these two types of theory.)

[4]For example, it is often the case that the strong coupling regime of one theory—in which problems are hard to solve by perturbative methods—can be translated into the weak coupling regime of its dual, in which problems are more tractable. See e.g. [27, p. 3].

[5]By this, we mean a Lorentzian manifold $\langle M, g_{ab} \rangle$, the metric field g_{ab} for which is a solution to the vacuum Einstein equations with negative cosmological constant, $G_{ab} + \Lambda g_{ab} = 0$ $(\Lambda < 0)$.

[6]One of the five consistent superstring theories – see e.g. [12, pp. 8ff.], and footnote 6 of §6.1.

[7]$\mathcal{N}$ denotes the number of supersymmetry generators in a QFT.

[8]As de Haro states [25, p. 3], generically, neither is the bulk pure AdS, nor is the boundary theory exactly conformal – it is sufficient that the bulk be asymptotically locally AdS and that the QFT on the boundary have a fixed point. (Recall that in QFT, renormalisation group flow may take us to fixed points, which are CFTs.) This is discussed further below.

[9]This being necessary for the formulation of the AdS/CFT correspondence.

element can be written in the form[10] [27, p. 15]

$$ds^2 = \frac{l^2}{r^2}\left(dr^2 + g_{ij}\left(r,x\right)dx^i dx^j\right), \tag{7.1.1}$$

where $g_{ij}\left(r,x\right)$ is an arbitrary function of the *radial coordinate*, r. The remaining coordinates x^i $(i = 1,\ldots,d := D-1)$ parameterise the boundary, which is of dimension d, and located at $r \to 0$.[11] The conformal metric at the boundary is $g_{(0)ij}\left(x\right) := g\left(0,x\right)$. Solving Einstein's equations amounts to finding $g_{ij}\left(r,x\right)$ given some initial data; because of the presence of a timelike boundary, choosing a spatial Cauchy surface at some initial time does not completely specify the problem [25, §2]. In addition, we must provide boundary conditions. $g_{ij}\left(r,x\right)$ has a regular expansion in a neighbourhood of $r = 0$ as

$$g_{ij}\left(r,x\right) = g_{(0)ij}\left(x\right) + rg_{(1)ij} + r^2 g_{(2)ij} + \ldots. \tag{7.1.2}$$

The coefficients in this expansion, apart from $g_{(0)ij}$ and $g_{(d)ij}$, are all determined algebraically from Einstein's equations. By contrast, the coefficients $g_{(0)ij}$ and $g_{(d)ij}$ are not so determined[12] – these are the boundary data.[13]

[10] Originally computed in [40].

[11] See [27, §2].

[12] Only the trace and divergence of $g_{(d)ij}$ is determined from Einstein's equations. [27, p. 15]

[13] We recover pure AdS space when $g_{(0)ij}\left(x\right)$ is chosen to be flat (cf. footnote 8). Though $g_{(0)ij}$ and $g_{(d)ij}$ are *a priori* arbitrary and unrelated, there do exist relations between the two quantities – see [27, p. 15]. For original sources, see [28, §2].

7.2 Background Independence

It is sometimes claimed (e.g. [80, p. 61]) that a holographic duality can be used to *define* a quantum gravity theory via the associated boundary theory.[14] Assuming this can be done, one must ask whether this quantum gravity theory manifests all necessary qualities of such a theory, potentially including background independence. This question—are holographic bulk theories background independent?—we address in §7.3. Before doing so, however, we discuss two definitions of background independence due to de Haro *et al.*, developed in the context of gauge/gravity dualities [25–27].

7.2.1 Minimalist Background Independence

De Haro *et al.* define *background independence in the minimalist sense* as follows: [27, §6.2]

Definition 13. (Background independence, minimalist)*: A theory is* background independence in the minimalist sense *if:*

(a) There is a generally covariant formulation of the dynamical laws of the theory that does not refer to any background metric (or background fields). 'Dynamical laws' is here understood in terms of an action and a corresponding path integral measure; and 'background' refers to fields whose values are not determined by corresponding equations of motion.

(b) The states and the quantities are invariant (or covariant, where appropriate) under diffeomorphisms, and also do not refer to any background metric.

Let us comment upon (a) and (b). Beginning with the former, de Haro *et al.* are explicit

[14] It is sometimes further claimed that the bulk theory *emerges* from the boundary theory – a notion subject to criticism in e.g. [33, 129].

that this criterion is to be understood as an extension of Def. 8 to quantum theories (see [27, p. 19], [30, p. 3]). This becomes clear, once one recognises (I) the requirement of the existence of a generally covariant action for the theory in question; (II) the prohibition of 'background fields' – which are intended to be understood as those fields which violate (i) and (ii) in Def. 8; and (since we are now dealing with quantum theories – see §5.2) (III) the inclusion of a requirement that the path integral measure not make reference to any background fields.

One worry does, however, exist regarding the overlap between 'background fields' in the sense of (a) in Def. 13, and fields which violate (i) and (ii) in Def. 8. As stated above, 'background field' refers only to "fields whose values are not determined by corresponding equations of motion" – and this coincides only with (i). Values of fields such as Θ^{abcd} in the case of **SR2** *are* "determined by corresponding equations of motion", so this field does *not* qualify as a 'background field', meaning that (a) *is* satisfied in the case of **SR2**. As we saw in §3.4.2, this is intuitively the incorrect result. This suggests that (a) in Def. 13 should be fixed to include a prohibition on 'unphysical fields', *à la* (ii) in Def. 8. Once such an amendment is made, (a) in Def. 13 *can* be understood as a reformulation of Def. 8 in the context of quantum theories.

Now consider (b); as stated at [27, p. 19], this condition is "novel". The first thing to note here is that "states and quantities" is understood to mean *physical* states and quantities – i.e., in the language of §2.1, those encoded exclusively in the elements of $\tilde{\mathcal{K}}$ of the theory in question.[15] The second thing to note about (b) is that it carries a tacit restriction to *relevant* diffeomorphisms [25, p. 16] – by which are meant those diffeomorphisms understood to be gauge transformations. With this in mind, (b) can be read as follows: 'Physical states and quantities must be invariant under diffeomorphisms understood to be gauge transformations.'

Having been re-posed thus, one may ask of (b): In what sense is this condition relevant to the notion of *background* in a physical theory? To answer this question, one must recognise that, for de Haro *et al.*, minimalist background independence is "a *necessary* requirement

[15]For a more precise definition of 'states and quantities', see [29, §3.2].

on any theory of quantum gravity which wishes to reproduce general relativity in the semi-classical limit" [30, p. 3]. Part of such a requirement is that physical quantities be invariant under those diffeomorphisms understood to be gauge transformations – thus justifying the inclusion of (b). Since, however, the notion of background independence discussed in this thesis regards the presence of 'background structure' in a theory, rather than compatibility with GR in the above sense, (b) can largely be set aside in our analyses.[16]

7.2.2 Extended Background Independence

In addition to Def. 13, de Haro *et al.* introduce *extended background independence*:[17]

> In the extended sense of [background independence], the initial or boundary conditions are also to be dynamically determined. This would be the case in a theory with either no boundary conditions at all, or boundary conditions which are determined by the dynamical equations.

Thus, extended background independence captures the desideratum that "boundary conditions should be obtained dynamically rather than by stipulation" [25, p. 7]. Extended background independence is manifestly in line with Smolin's requirement that all 'background structures' ultimately be excised from our physics (see e.g. [122, p. 204]).

[16] Though, for completeness, we *do* assess whether (b) is satisfied by bulk theories in holographic dualities. Note also that this is not to question the value of the programme of de Haro *et al.*

[17] This quote from [30, p. 4]; see also [25, §2].

7.3 Background Independence and Holography

7.3.1 Minimalist Background Independence

Do bulk theories in AdS/CFT dualities satisfy Def. 13? At the classical level, we have seen in §7.1 that the dynamical equations of these theories are Einstein's equations with negative cosmological constant; these can be derived from a generally covariant action. Moreover, as de Haro states, "adding (string-theoretical) classical matter fields does not change this picture: since these contribute covariant terms to the equations of motion" [25, p. 5]. Indeed, even adding quantum corrections does not alter this verdict, for "In the full quantum version of AdS/CFT, quantum corrections manifest themselves as higher-order corrections in powers of the curvature to the classical action. These are generally covariant as well." Since none of these correction terms involve fields which violate (i) and (ii) in Def. 8, one might conclude that such bulk theories *do* satisfy (a) of Def. 13.[18]

This being said, one must recall here subtleties regarding theories with boundary conditions, discussed in §3.4.2 in the context of Def. 8. First, let the KPMs of the bulk theories (for simplicity, considered here at the classical level) be the KPMs of vacuum GR, with negative cosmological constant. Let the DPMs be those KPMs which satisfy Einstein's equations; and let the choice of boundary conditions—fixed by $g_{(0)ij}$ and $g_{(d)ij}$—fix a $\mathcal{B} \subset \mathcal{K}$.

At this juncture we face three interpretive options, depending upon how a *theory* is defined. First, one may state that a theory plus boundary conditions is not itself a new theory; background independence of a theory plus boundary conditions should be assessed with respect to $\mathcal{D} \subset \mathcal{K}$. Second, one may consider the $\mathcal{B}$ picked out by the boundary conditions to specify the KPMs of a new theory. Then, background independence should be assessed with respect to whether the dynamical laws which pick out $\mathcal{D} \cap \mathcal{B} \subset \mathcal{B}$ can be obtained as per

[18] In addition, these corrections do not lead to violations of Def. 13 at the level of the path integral measure.

Def. 8. Third, one may impose boundary conditions at the level of dynamics; background independence should then be assessed with respect to whether the dynamical laws which pick out $\mathcal{D} \cap \mathcal{B} \subset \mathcal{K}$ can be obtained as per Def. 8.[19]

In the first of these cases, since GR is background independent on Def. 8, it is clear that the bulk theories in question are also background independent in this sense, and so satisfy (a) of Def. 13. As we saw in §3.4.2, in the second case, the definition of background independence is *also* satisfied by these theories with boundary conditions, for the very same action principle suffices to pick out $\mathcal{D} \cap \mathcal{B}$, when KPMs are restricted to $\mathcal{B}$. On the third conception, however, it is *prima facie* unclear whether such theories will satisfy Def. 8, and so (a) of Def. 13, for it is unclear whether there exists an action principle which picks out $\mathcal{D} \cap \mathcal{B} \subset \mathcal{K}$. While de Haro argues that, since boundary conditions must be imposed[20] in *every* element of $\mathcal{D}$ for the bulk theories in question, it is not correct to regard each choice of boundary conditions as picking out a different theory (thereby favouring the first interpretive route above) [30, pp. 2, 9], what we wish to highlight here is that there does exist room in logical space to make an alternative move, on which Def. 8, and so (a) of Def. 13, may be violated.

In any case, turn now to (b) in Def. 13. Following [25, §2.3.3], distinguish:

(a1) Diffeomorphisms which preserve the asymptotic form of the metric;

(a2) Diffeomorphisms which change the conformal class of the asymptotic metric.

Furthermore, within (a1), distinguish:

(a1)-1 Diffeomorphisms which preserve the *boundary metric* (i.e. the representative of the conformal class).

[19] See §3.4.2.

[20] I.e. $g_{(0)ij}$ and $g_{(d)ij}$ must be specified.

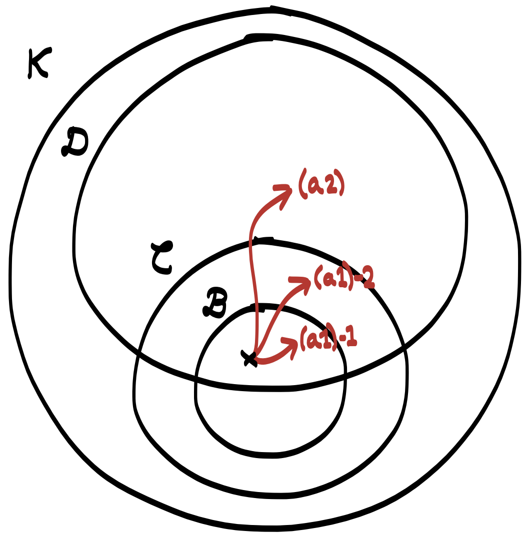

Figure 7.1: The action of diffeomorphisms of types (a1)-1, (a1)-2, and (a2) in the space of models of a bulk theory in an AdS/CFT duality.

(a1)-2 Diffeomorphisms which do not preserve the *boundary metric* (i.e. they do not preserve the *representative* of the conformal class, though they preserve the conformal class) – these are conformal transformations.

Let us attempt to visualise these different types of diffeomorphism. As we have seen, "in order to obtain a solution of the equations of motion of general relativity in spaces with a negative cosmological constant, boundary conditions need to be specified" [30, p. 2]. Hence, for the case in which the theory in question is GR with negative cosmological constant, to pick out any point in $\mathcal{D}$, one must specify boundary conditions. Now, choose one such point in $\mathcal{D}$. Then, call the class of models which obey the same boundary conditions $\mathcal{B}$, and the class of models which have the same conformal class of boundary conditions $\mathcal{C}$. The picture that emerges is as per Fig. 7.1.[21] On this setup, transformations of type (a1)-1 keep us in $\mathcal{B}$; generic transformations of type (a1)-2 take us to a point out of $\mathcal{B}$ but still in $\mathcal{C}$; and transformations of type (a2) take us to a point out of $\mathcal{B}$ to a point in $\mathcal{D}$ and not in $\mathcal{C}$.

[21] Here, by extension from Fig. 2.1, we draw $\mathcal{B}$ and $\mathcal{C}$ overlapping the border of $\mathcal{D}$, to allow for the possibility that the fields of the theory may obey some other dynamics yet nevertheless satisfy the same boundary conditions – though this possibility will not be relevant in what follows.

While de Haro adopts a standard view of diffeomorphisms of type (a1)-1 as relating gauge-equivalent models,[22] he also endorses a line according to which diffeomorphisms of type (a2) relate physically *inequivalent* models [25, p. 6]. The situation regarding diffeomorphisms of type (a1)-2 is more subtle. As discussed at [25, §2.3.3], for odd d, bulk theories manifest asymptotic conformal invariance; as a result, physical quantities are invariant under transformations of type (a1)-2. For even d, however, "the asymptotic conformal invariance is broken by the regularisation of [a] large-volume divergence. ... As a consequence, physical quantities such as the stress-energy tensor ... no longer transform covariantly, but pick up an anomalous term" [25, p. 7]. This is the *diffeomorphism anomaly*.

What should one make of these results? First, if, in line with De Haro, one regards diffeomorphisms of type (a2) as not being gauge transformations, then it is clear that these transformations are not *relevant* in the sense of §7.3.1. Therefore, the fact that physical states and quantities are not preserved under such transformations does not lead to a violation of (b). Similarly, if diffeomorphisms of type (a1)-2 in the case of even d are not regarded as gauge transformations, then they are also excluded from consideration in assessments of whether (b) in Def. 13 is satisfied – so the fact that physical states and quantities are not preserved under such transformations again presents no violation of this condition. This, indeed, is the preferred move of de Haro *et al.* [25,27,29,30].[23,24] Thus, one *can* tell a plausible story regarding the satisfaction of Def. 13 by bulk theories in AdS/CFT dualities.[25]

Finally, it is worth commenting on whether bulk theories in AdS/CFT dualities satisfy extended background independence, as presented in §7.2.2. De Haro points out that the straightforward answer here, as with GR, is *no*, for boundary conditions in such theories are *not* dynamically determined [25, §2.3.4]. As de Haro argues, however, this result should not be

[22] *Pace* Belot – see §2.2.

[23] Note that such a move is not possible on the interpretational approach to symmetries, presented in §2.1.

[24] One might (legitimately) worry, however, that de Haro's making this move *post facto* is too easy, and that some prior demarcation of which transformations qualify as physical is required.

[25] Though, of course, if one *does* maintain that diffeomorphisms of type (a1)-2 in the case of even d are gauge transformations, then (b) in Def. 13 is violated.

construed as problematic, for extended background independence is best considered a heuristic *desideratum* on a quantum gravity theory, rather than as a necessary condition [25, §2.3.4].[26]

7.3.2 Other Approaches

Above, we have seen that there is a plausible route to arguing that Def. 13 is satisfied by bulk theories in AdS/CFT dualities – and since, as stated in §7.2.1, (a) in Def. 13 is understood as a reformulation of Def. 8 in the sense of §5.2, Def. 8 is also satisfied by these theories. This being the case, turn now to Def. 5, Def. 6, and Def. 10. On Def. 5 and Def. 6, it is clear that evaluations of bulk theories with respect to these definitions must be made on a case-by-case basis. Nevertheless, none of the best-known bulk theories in AdS/CFT dualities—understood as classical theories plus quantum corrections—involve fixed fields;[27] this is good *prima facie* evidence that Def. 6 is satisfied by these theories.[28]

Turn now to Def. 5. Here, the bulk theories under consideration do not satisfy this definition, for these theories can be understood as vacuum GR with negative cosmological constant, plus corrections. However, recall from §3.3.3 that GR itself was found to have an absolute object – the square root of the metric determinant; given this, one naturally also expects these bulk theories to manifest an absolute object. This, indeed, is exactly the same point that was made in the context of perturbative string theory, in §6.3.

Finally, consider Belot's account of background independence. To establish which of the definitions in §3.5.2 is satisfied by bulk theories in AdS/CFT dualities, first recall that Belot understands boundary conditions to be part of the *definition* of a theory;[29] this permits him to

[26] In this thesis, unlike de Haro, we leave open whether *any* account of background independence is strictly a necessary condition on a quantum gravity theory.

[27] See e.g. [3].

[28] Moreover, since these theories can be understood as GR plus corrections, there also exists reason to suppose that alternative formulations of these theories featuring fixed fields do *not* exist – cf. §3.4.1.

[29] Accordingly, Belot can be understood as following either the second or third interpretative branch above.

disregard diffeomorphisms which do not preserve specified boundary conditions (e.g. those of type (a1)-1 and (a2)). Then, Belot states that non-trivial diffeomorphisms of type (a1)-1 do *not* relate physically equivalent models. Making the analogous interpretation in the case of bulk theories, we see that such theories should also qualify as *nearly background independent*, for the quotient of the group of all diffeomorphisms by the group of diffeomorphisms which act trivially on the boundary conditions is again finite-dimensional. On the other hand, if one follows de Haro in regarding *all* diffeomorphisms of type (a1)-1 as relating physically equivalent models, then one will—by analogy with the discussion in §3.5.3—consider bulk theories to be *fully* background independent.

7.4 Coda

Having assessed the background independence of bulk theories in AdS/CFT dualities, we close by considering some further issues related to the above discussion. In §7.4.1, we return to the issue of the definition of general covariance (cf. §3.1), in light of comments made at [25, p. 7]. In §7.4.2, we assess whether one can exploit holographic dualities to render background independent theories which are *prima facie* background dependent.

7.4.1 General Covariance

At [25, p. 7], de Haro *et al.* claim that the presence of the diffeomorphism anomaly "contradicts the claim, often seen in discussions of background independence (and going back to Kretschmann's objection to Einstein's claims that general covariance selected general relativity), that 'any' theory can be given a manifestly covariant formulation" (see also [27, p. 21]). This is questionable, as it appears to conflate the trivial notion of general covariance discussed in §3.1, with more substantive notions, such as diffeomorphism invariance (see §3.2).

Since a bulk theory typically will be formulable in the coordinate-free language of differential geometry (notwithstanding the diffeomorphism anomaly), such a theory *can* be made generally covariant in the trivial sense. On the other hand, it is true that the diffeomorphism anomaly breaks the diffeomorphism invariance of the theory. It is wrong, however, to conflate these two senses of 'general covariance', and thereby to conclude that the diffeomorphism anomaly provides a counterexample to the Kretschmann point.[30]

7.4.2 Inter-Theory Translation and Background Independence

As a second point of interest, let us reflect on the possibility of inter-theory translation between the CFT and the bulk in an AdS/CFT duality. Suppose we accept that the bulk theory in such a duality is background independent. As de Haro states, the CFT, by contrast, violates Def. 13 of background independence, not least because in this theory "the metric field is not determined by the equations of motion" [25, p. 7] – thereby violating (a).

Given this, we appear to have in our possession a duality according to which one of the theories is background independent, and the other is not. In itself, this is an interesting result. One might, however, go further, by claiming that the CFT can be *rendered* background independent via holography (see e.g. [117, p. 7]).[31] Here, however, we urge caution: the two dual theories are *a priori* very different, and (absent an overarching theory in which degrees of freedom of both can be defined, or some story regarding how just *one* of the theories is metaphysically privileged – see e.g. [105]) should be evaluated *on their own terms*.[32] If one follows this mantra, then one *cannot* straightforwardly regard the CFT as being rendered background independent via the holographic duality.

[30] For related discussion regarding conflating these two different senses of general covariance, see [99].

[31] In a sense, this is the reverse of the claim often found in the literature that a quantum gravity theory—here, the bulk theory—*emerges* from the CFT, via the holographic duality. For critical discussion, see e.g. [129].

[32] C.f. §§4.2.3, 4.3.2.

Chapter 8

Conclusions

In this thesis, we have presented four definitions of background independence: Def. 5 (in terms of the absence of absolute objects), Def. 6 (in terms of the non-existence of theory formulations featuring fixed fields), Def. 8 (in terms of variational principles), and Def. 10 (in terms of a coincidence of geometrical and physical degrees of freedom in a theory). Although all such approaches to background independence face difficulties (§3), we have followed a pluralistic approach, applying such definitions in parallel in order to attain verdicts on the background independence of various spacetime theories.

In the classical case (§4), we found that while NG and NCT are background dependent, there exists a majority verdict that TPG and KKT are background independent (modulo subtleties); moreover, we argued that such theories should *not* simply be understood as reformulations of GR. Thus, background independence cannot be a *defining* feature of GR. In the quantum case (§§6, 7), we found that the majority verdict is that perturbative string theories and bulk holographic theories *are* background independent (again, modulo subtleties). Thus, even those who subscribe to background independence being a necessary quality of a quantum gravity theory should not be deterred from the study of these theories.

Possible extensions abound. On the one hand, definitions of background independence such as that due to Gryb [52] have been overlooked. On the other hand, there exist several classical and quantum gravity theories—most notably *shape dynamics* [8] and *loop quantum gravity* [114,116]—for which a detailed analysis in the style of the above has yet to be pursued. Since many of the claims to the effect that background independence is a necessary quality of a quantum gravity theory issue from the progenitors of such theories,[1] such work is of manifest importance. Our hope is that the systematic, precise nature of the above work be a guide for future research in this field.

[1]E.g. [113, 114, 123, 124].

Acknowledgements

I am profoundly grateful to Oliver Pooley for his meticulous supervision throughout the development of this thesis. I am also indebted to Harvey Brown, Sebastian de Haro, Nick Huggett, Keizo Matsubara, Dean Rickles, and Simon Saunders, for invaluable discussions regarding background independence. My B.Phil. studies were supported by an AHRC grant at the University of Oxford; I am also grateful to Merton College for support; and to the Space and Time After Quantum Gravity project for a visiting scholarship to the University of Illinois at Chicago in April 2016, where some of this work was completed.

Bibliography

[1] Ruben Aldrovandi, and José Geraldo Pereira, *Teleparallel Gravity: An Introduction*, Berlin: Springer, 2013.

[2] Luis Alvarez-Gaumé and Edward Witten, "Gravitational Anomalies", Nuclear Physics B 234, pp. 269-330, 1983.

[3] Martin Ammon and Johanna Erdmenger, *Gauge/Gravity Duality: Foundations and Applications*, Cambridge: Cambridge University Press, 2015.

[4] J. L. Anderson, "Relativity Principles and the Role of Coordinates in Physics", in H.-Y. Chiu and W. F. Hoffmann (eds.), *Gravitation and Relativity*, New York: W. A. Benjamin, pp. 175-194, 1964.

[5] J. L. Anderson, *Principles of Relativity Physics*, New York: Academic Press, 1967.

[6] Lars Andersson, "Momenta and Reduction in General Relativity", Journal of Geometry and Physics 4, pp. 289-314, 1987.

[7] V. de Andrade, L. Guillen, and J. Pereira, "Teleparallel Gravity: An Overview", arXiv:gr-qc/0011087, 2000.

[8] Julian Barbour, "Shape Dynamics: An Introduction", in *Proceedings of the Conference Quantum Field Theory and Gravity*, Regensburg, 2010.

[9] Julian Barbour, “Does Time Differ From Change? Philosophical Appraisal of the Problem of Time in Quantum Gravity and in Physics: A response”, Studies in the History and Philosophy of Modern Physics 52, pp. 55-61, 2015.

[10] David Bailin and Alex Love, “Kaluza-Klein Theories”, Reports on Progress in Physics 50, pp. 1087-1170, 1987.

[11] Jonathan Bain, “Theories of Newtonian Gravity and Empirical Indistinguishability”, Studies in the History and Philosophy of Modern Physics 35, pp. 345-376, 2004.

[12] Katrin Becker, Melanie Becker, and John Schwarz, *String Theory and M-Theory: A Modern Introduction*, Cambridge: Cambridge University Press, 2007.

[13] Gordon Belot, “An Elementary Notion of Gauge Equivalence”, General Relativity and Gravitation 40, pp. 199-215, 2008.

[14] Gordon Belot, “Background-Independence”, General Relativity and Gravitation 43, pp. 2865-2884, 2011.

[15] Gordon Belot, “Symmetry and Equivalence”, in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, pp. 318-339, 2013.

[16] Gordon Belot, “Fifty Million Elvis Fans Can’t be Wrong”, 2016. Available at: https://sites.google.com/site/gordonbelot/home/papers-etc.

[17] Ralph Blumenhagen, Dieter Lüst, and Stefan Theisen, *Basic Concepts of String Theory*, Springer Series in Theoretical and Mathematical Physics, Berlin: Springer, 2013.

[18] Harvey R. Brown, “On the Role of Special Relativity in General Relativity”, International Studies in the Philosophy of Science 11, pp. 67-81, 1997.

[19] Harvey R. Brown, *Physical Relativity: Space-Time Structure From a Dynamical Perspective*, Oxford: Oxford University Press, 2005.

[20] Harvey R. Brown and Dennis Lehmkuhl, "Einstein, the Reality of Space, and the Action-Reaction Principle", in Partha Ghose (ed.), *The Nature of Reality*, Pickering and Chatto, 2015.

[21] Harvey R. Brown, Dennis Lehmkuhl, and James Read, "On the Local Validity of Special Relativity", 2016. (In preparation.)

[22] 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.

[23] Joy Christian, "Exactly Soluble Sector of Quantum Gravity", Physical Review D 56(8), pp. 4844-4877, 1997.

[24] Piet Claus, Renata Kallosh, Antoine Van Proeyen, "M5-Brane and Superconformal $(0, 2)$ Tensor Multiplet in Six Dimensions", Nuclear Physics B 518, pp. 117-150, 1998.

[25] Sebastian de Haro, "Dualities and Emergent Gravity: Gauge/Gravity Duality", forthcoming in Studies in History and Philosophy of Modern Physics, 2016.

[26] Sebastian de Haro, "Invisibility of Diffeomorphisms", 2016.

[27] Sebastian de Haro, Daniel R. Mayerson and Jeremy Butterfield, "Conceptual Aspects of Gauge/Gravity Duality", forthcoming in Foundations of Physics, 2016.

[28] Sebastian de Haro, Kostas Skenderis and Sergey N. Solodukhin, "Holographic Reconstruction of Spacetime and Renormalization in the AdS/CFT Correspondence", Communications in Mathematical Physics 217, pp. 595-622, 2001.

[29] Sebastian de Haro, Nicholas Teh and Jeremy Butterfield, "Comparing Dualities and Gauge Symmetries", forthcoming in Studies in History and Philosophy of Modern Physics, 2016.

[30] Sebastian de Haro, "Reply to James Read on Background-Independence", 2016.

[31] Neil Dewar, “Symmetries and the Philosophy of Language”, Studies in the History and Philosophy of Modern Physics 52, pp. 317-327, 2015.

[32] Neil Dewar, “Maxwell-Cartan Gravitation”, 2016. Available at: http://philsci-archive.pitt.edu/12045/.

[33] 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.

[34] John Earman, *World Enough and Space-Time: Absolute Versus Relational Theories of Space and Time*, Cambridge, MA: MIT Press, 1989.

[35] John Earman, “The Cosmological Constant, the Fate of the Universe, Unimodular Gravity, and All That” Studies in History and Philosophy of Science 34, pp. 559-577, 2003.

[36] John Earman, “The Implications of General Covariance for the Ontology and Ideology of Spacetime”, in Dennis Dieks (ed.), *The Ontology of Spacetime*, Volume 1 of *Philosophy and Foundations of Physics*, Amsterdam: Elsevier, pp. 3-24, 2006.

[37] John Earman, “Two Challenges to the Requirement of Substantive General Covariance”, Synthese 148(2), pp. 443-468, 2006.

[38] 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.

[39] J. Ehlers, Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie, in J. Nitsch, J. Pfarr, and E. W. Stachow (eds.), *Grundlagen Probleme der Modernen Physik*, Wissenschaftsverlag, 1981.

[40] C. Fefferman and C. Robin Graham, “Conformal Invariants”, in *Elie Cartan et les Mathématiques d’aujourd’hui*, pp. 95ff., Astérisque, 1985.

[41] J. A. Ferrari, J. Griego and E. E. Falco, “The Kaluza-Klein Theory and Four-Dimensional Spacetime”, General Relativity and Gravitation 21, pp. 69-78, 1989.

[42] M. Ferraris, M. Francaviglia and C. Reina, “A Constructive Approach to Bundles of Geometric Objects on a Differentiable Manifold”, Journal of Mathematical Physics 24:120, 1983.

[43] Richard P. Feynman, *The Character of Physical Law*, Cambridge MA: MIT Press, 1965.

[44] Richard P. Feynman, Fernando B. Morinigo and William G. Wagner, *Feynman Lectures on Gravitation*, Reading, MA: Addison-Wesley, 1995.

[45] Michael Friedman, “Relativity Principles, Absolute Objects and Symmetry Groups”, in P. Suppes (ed.), *Space, Time, and Geometry*, pp. 296-320, Dordrecht: D. Reidelm, 1973.

[46] Michael Friedman, *Foundations of Space-Time Theories: Relativistic Physics and the Philosophy of Science*, Princeton: Princeton University Press, 1983.

[47] M. Gaul and C. Rovelli, “Loop Quantum Gravity and the Meaning of Diffeomorphism Invariance”, in J. Kowalski-Glikman (ed.), *Towards Quantum Gravity: Proceeding of the XXXV International Winter School on Theoretical Physics Held in Polanica, Poland, 2-11 February 1999*, Volume 541 of Lecture Notes in Physics, Berlin, Heidelberg: Springer, pp. 277-324, 2000.

[48] G. W. Gibbons and J. B. Hartle, “Real Tunnelling Geometries and the Large-Scale Topology of the Universe”, Physical Review D 42, pp. 2458-2468, 1990.

[49] D. Giulini, “Remarks on the Notions of General Covariance and Background Independence”, in I.-O. Stamatescu and E. Seiler (eds.), *Lecture Notes in Physics*, Volume 721, pp. 105-120. Berlin, Heidelberg: Springer, 2007.

[50] Hilary Greaves and David Wallace, “Empirical Consequences of Symmetries”, British Journal for the Philosophy of Science 65, pp. 59-89, 2014.

[51] Michael B. Green, John H. Schwarz and Edward Witten, *Superstring Theory*, Vol. 1, Cambridge: Cambridge University Press, 1987.

[52] Sean Gryb, "A Definition of Background Independence", Classical and Quantum Gravity 27(2), 215018, 2010.

[53] James Hartle and Stephen Hawking, "Wavefunction of the Universe", Physical Review D 28(12), pp. 2960-2975, 1983.

[54] M. Henningson and K. Skenderis, "The Holographic Weyl Anomaly", Journal of High Energy Physics 9807(23), 1998.

[55] Carl Hoefer, "Energy Conservation in GTR", in Studies in the History and Philosophy of Modern Physics 31(2), 2000.

[56] Gerard 't Hooft, "Dimensional Reduction in Quantum Gravity", arXiv:gr-qc/9310026, 1993.

[57] Gary T. Horowitz, "String Theory Without Spacetime", in Abhay Ashtekar and John Stachel (eds.), *Conceptual Problems in Quantum Gravity*, Springer, 1988.

[58] Gary T. Horowitz and Joseph Polchinski, "Gauge/Gravity Duality", in *Approaches to Quantum Gravity: Toward a New Understanding of Space, Time and Matter*, D. Oriti, Cambridge University Press, 2009.

[59] Nick Huggett, "Target Space $\neq$ Space", forthcoming in Studies in the History and Philosophy of Modern Physics, 2016.

[60] Nick Huggett, "The String Theoretic Explanation of General Relativity", 2016.

[61] Nick Huggett and Tiziana Vistarini, "Deriving General Relativity from String Theory", Philosophy of Science 82(5), pp. 1163-1174, 2015.

[62] Jenann Ismael and Bas C. van Fraassen, "Symmetries as a Guide to Superfluous Theoretical Structure", in Katherine Brading and Elena Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge University Press, 2003.

[63] John David Jackson, *Classical Electrodynamics*, third edition, John Wiley & Sons, Inc., 1999.

[64] Theodor Kaluza, "Zum unitätsproblem in der physik", Sitzungsber. Preuss. Akad. Wiss. Berlin. Math. Phys. 33, pp. 966-972, 1921.

[65] B. Khesin and R. Wendt, *The Geometry of Infinite-Dimensional Groups*, Berlin: Springer, 2009.

[66] Oskar Klein, "Quantentheorie und fünfdimensionale relativitätstheorie", Zeitschrift für physik A 37, pp. 895-906, 1926.

[67] 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.

[68] Eleanor Knox, "Newtonian Spacetime Structure in Light of the Equivalence Principle", British Journal for the Philosophy of Science 65(4), pp. 863-880, 2014.

[69] Erich Kretschmann, "Über den physikalischen Sinn der Relativitätspostulate", Annalen der Physik 53, pp. 575-614, 1917.

[70] Karel Kuchař, "Canonical Quantisation of Gravity", in W. Israel (ed.), *Relativity, Astrophysics, and Cosmology*, D. Reidel, Dordrecht, pp. 237-288, 1973.

[71] Karel Kuchař, "Canonical Methods of Quantisation", in C. J. Isham, R. Penrose, and D. W. Sciama (eds.), *Quantum Gravity 2: A Second Oxford Symposium*, Oxford: Clarendon Press, pp. 329ff., 1981.

[72] H. Künzle, "Galilei and Lorentz Structures on Space-Time: Comparison of the Corresponding Geometry and Physics" Annales Institute Henri Poincaré 17, pp. 337-362, 1972.

[73] H. Künzle, "Covariant Newtonian Limts of Lorentz Space-Times", General Relativity and Gravitation 7, pp. 445-457, 1976.

[74] Vincent Lam, "Gravitational and Nongravitational Energy: The Need for Background Structures", Philosophy of Science 78(5), 2011.

[75] D. L. Lee, A. P. Lightman and W.-T. Ni, "Conservation Laws and Variational Principles in Metric Theories of Gravity", Physical Review D 10, pp. 1685-1700, 1974.

[76] David Lovelock, "The Einstein Tensor and its Generalisations", Journal of Mathematical Physics 12, pp. 498-501, 1971.

[77] Anna Maidens, "Symmetry Groups, Absolute Objects and Action Principles in General Relativity", Studies in the History and Philosophy of Modern Physics 29, pp. 245-272, 1998.

[78] David Malament, *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*, Chicago: University of Chicago Press, 2012.

[79] Juan Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", Advances in Theoretical and Mathematical Physics 2, pp. 231-252, 1998.

[80] Juan Maldacena, "The Illusion of Gravity", Scientific American, pp. 56-63, November 2005.

[81] Flavio Mercati, "A Shape Dynamics Tutorial", arXiv:1409.0105, 2014.

[82] Charles Misner, Kip Thorne, and John Wheeler, *Gravitation*, San Francisco: Freeman & Co., 1973.

[83] Thomas Møller-Nielsen, "Invariance, Interpretation, and Motivation", 2016.

[84] John Norton, "Coordinates and Covariance: Einstein's View of Space-Time and the Modern View", Foundations of Physics 19, pp. 1215-1263, 1989.

[85] John Norton, "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute", Reports on Progress in Physics 56(7), pp. 791-858, 1993.

[86] V. I. Ogievetsky and I. V. Polubarinov, "Spinors in Gravitation Theory", Soviet Physics, Journal of Experimental and Theoretical Physics 21, pp. 1093ff. (Russian volume 48, pp. 1625ff.), 1965.

[87] Hans Ohanian and Remo Ruffini, *Gravitation and Spacetime*, third edition, Cambridge: Cambridge University Press, 2013.

[88] J. M. Overduin and P. S. Wesson, "Kaluza-Klein Gravity", Physics Reports 283, pp. 303-378, 1997.

[89] T. Padmanabhan, "From Gravitons to Gravity: Myth and Reality", International Journal of Modern Physics D 17, pp. 367-398, 2008.

[90] Antonello Pasini, "A Conceptual Introduction to Kaluza-Klein Theory", European Journal of Physics 9, pp. 289-296, 1988.

[91] André Morgado Patrício, "5D Kaluza-Klein Theories - A Brief Review", 2013. Available at: https://fenix.tecnico.ulisboa.pt/downloadFile/3779580604342/kaluza.pdf.

[92] Roger Penrose, *The Road to Reality*, London: Jonathan Cape, 2004.

[93] J. G. Pereira, "Teleparallelism: A New Insight Into Gravity", in Abhay Ashtekar and Vesselin Petkov (eds.), *Handbook of Spacetime*, Berlin: Springer, 2014.

[94] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, Westview Press, 1995.

[95] 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.

[96] J. Brian Pitts, “Empirical Equivalence, Artificial Gauge Freedom and a Generalized Kretschmann Objection”, arXiv:0911.5400, 2009.

[97] J. Brian Pitts, “Gauge-Invariant Localization of Infinitely Many Gravitational Energies from All Possible Auxiliary Structures”, General Relativity and Gravitation 42, 2010.

[98] J. Brian Pitts, “The Nontriviality of Trivial General Covariance: How Electrons Restrict ‘Time’ Coordinates, Spinors (Almost) Fit into Tensor Calculus, and $\frac{7}{16}$ of a Tetrad is Surplus Structure”, Studies in History and Philosophy of Modern Physics 43, pp. 1-24, 2012.

[99] J. Brian Pitts and W. C. Schieve, “Universally Coupled Massive Gravity”, Theoretical and Mathematical Physics 151(2), pp. 700-717, 2007.

[100] Joseph Polchinski, *String Theory*, Vol. 1, Cambridge: Cambridge University Press, 1998.

[101] Joseph Polchinski, “Dualities of Fields and Strings”, forthcoming in Studies in the History and Philosophy of Modern Physics, 2016.

[102] Oliver Pooley, “Substantive General Covariance: Another Decade of Dispute”, in Mauricio Suárez, Mauro Dorato and Miklós Rédei (eds.), *EPSA Philosophical Issues in the Sciences: Launch of the European Philosophy of Science Association*, Springer, pp. 197-209, 2010.

[103] Oliver Pooley, “Substantivalist and Relationist Approaches to Spacetime”, in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, 2013.

[104] Oliver Pooley, “Background Independence, Diffeomorphism Invariance, and the Meaning of Coordinates”, in Dennis Lehmkuhl (ed.), *Towards a Theory of Spacetime Theories*, Birkhäuser. (Forthcoming.)

[105] James Read, “The Interpretation of String-Theoretic Dualities”, Foundations of Physics 46(2), pp. 209-235, 2016.

[106] Dean Rickles, “Quantum Gravity: A Primer for Philosophers”, in *The Ashgate Companion to Contemporary Philosophy of Physics*, Ashgate, 2008.

[107] Dean Rickles, “Who’s Afraid of Background Independence?”, in D. Dieks (ed.), *The Ontology of Spacetime II*, Volume 4 of Philosophy and Foundations of Physics, pp. 133-152. Amsterdam: Elsevier, 2008.

[108] Dean Rickles, “A Philosopher Looks at String Dualities”, Studies in History and Philosophy of Modern Physics 42, pp. 54-67, 2011.

[109] Dean Rickles, “AdS/CFT Duality and the Emergence of Spacetime”, Studies in History and Philosophy of Modern Physics 44, pp. 312-320, 2013.

[110] Dean Rickles, “Mirror Symmetry and Other Miracles in Superstring Theory”, Foundations of Physics 43, pp. 54-80, 2013.

[111] N. Rosen, “Flat Space and Variational Principle”, in B. Hoffmann (ed.), *Perspectives in Geometry and Relativity: Essays in Honor of Václav Hlavatý*, Bloomington: Indiana University Press, Chapter 33, pp. 325-327, 1966.

[112] Biel Cardona Rotger, “Elementary Features of Kaluza-Klein Theories”, 2014. Available at: http://diposit.ub.edu/dspace/bitstream/2445/59761/1/TFG_Cardona_Rotger_Biel.pdf.

[113] Carlo Rovelli, “Quantum Spacetime: What Do We Know?”, in Craig Callender and Nick Huggett (eds.), *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*, Cambridge: Cambridge University Press, pp. 101-122, 2001.

[114] Carlo Rovelli, *Quantum Gravity*, Cambridge: Cambridge University Press, 2004.

[115] Carlo Rovelli, "A Critical Look at Strings", Foundations of Physics 43, pp. 8-20, 2013.

[116] Carlo Rovelli and Francesca Vidotto, *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*, Cambridge: Cambridge University Press, 2014.

[117] Moshe Rozali, "Comments on Background Independence and Gauge Redundancies", Advanced Science Letters 2(2), pp. 244-250, 2009.

[118] Ryan Samaroo, "On Identifying Background-Structure in Classical Field Theories", Philosophy of Science 78, pp. 1070-1081, 2011.

[119] Simon Saunders, "Physics and Leibniz's Principles", in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, 2003.

[120] Simon Saunders, "Mirroring as an A Priori Symmetry", Philosophy of Science 74(4), pp. 452-480, 2007.

[121] Nathan Seiberg, "Emergent Spacetime", in David Gross, Marc Henneaux and Alexander Sevrin (eds.), *The Quantum Structure of Space and Time*, Proceedings of the 23rd Solvay Conference on Physics, World Scientific, 2007.

[122] Lee Smolin, "Time, Structure and Evolution in Cosmology", in A. Ashtekar, R. S. Cohen, D. Howard, J. Renn, S. Sarkar, and A. Shimony (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*, vol. 234 of *Boston Studies in the Philosophy of Science*, pp. 221-274, Dordrecht: Kluwer, 2003.

[123] Lee Smolin, "The Case for Background Independence", in D. Rickles, S. French, and J. Saatsi (eds.), *The Structural Foundations of Quantum Gravity*, Oxford: Oxford University Press, pp. 196-239, 2006.

[124] Lee Smolin, *The Trouble with Physics: The Rise of String Theory, The Fall of a Science and What Comes Next*, Penguin, 2008.

[125] Rafael Sorkin, "An Example Relevant to the Kretschmann-Einstein Debate", Modern Physics Letters A 17, pp. 695-700, 2002.

[126] Patrick Suppes, "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences", Synthese 12, 1960.

[127] Adán Sus, *General Relativity and the Physical Content of General Covariance*, PhD Thesis, Universitat Autònoma de Barcelona, 2008.

[128] Adán Sus, "Absolute Objects and General Relativity: Dynamical Considerations", in M. Suarez, M. Dorato, and M. Rédei (eds.), *EPSA Philosophical Issues in the Sciences: Launch of the European Philosophy of Science Association*, Volume 2, Chapter 23, Springer, pp. 239-249, 2010.

[129] Nicholas Teh, "Holography and Emergence", in Studies in History and Philosophy of Modern Physics 44, pp. 300-311, 2013.

[130] 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.

[131] Roberto Torretti, "Space-Time Physics and the Philosophy of Science: Review of *Foundations of Space-Time Theories: Relativistic Physics and the Philosophy of Science* by Michael Friedman", British Journal for the Philosophy of Science 35, pp. 280-292, 1984.

[132] 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.

[133] Cumrun Vafa, "Geometric Physics", in *Proceedings of the International Congress of Mathematics, Vol. 1*, G. Fischer and U. Rehmann (eds.), 1998.

[134] Bas C. van Fraassen, *The Scientific Image*, Oxford: Oxford University Press, 1980.

[135] Bas C. van Fraassen, *Laws and Symmetry*, Oxford: Oxford University Press, 1989.

[136] Antonio Vassallo, "Can Bohmian Mechanics be Made Background Independent?", Studies in History and Philosophy of Modern Physics 52, pp. 242-250, 2015.

[137] Antonio Vassallo, "General Covariance, Diffeomorphism Invariance, and Background Independence in 5 Dimensions", in Tomasz Bigaj and Christian Wüthrich (eds.), *Metaphysics in Contemporary Physics*, Amsterdam: Rodopi. (Forthcoming)

[138] Robert Wald, *General Relativity*, Chicago: University of Chicago Press, 1984.

[139] David Wallace, "Fields as Bodies: A Unified Presentation of Spacetime and Internal Gauge Symmetry", 2015. Available at: http://philsci-archive.pitt.edu/11337/.

[140] David Wallace, "Who's Afraid of Coordinate Systems? An Essay on Representation of Spacetime Structure", 2016. Available at: http://philsci-archive.pitt.edu/11988/.

[141] James Owen Weatherall, "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?", forthcoming in Erkenntnis, 2016.

[142] James Owen Weatherall, "Understanding Gauge", forthcoming in Philosophy of Science, 2016.

[143] James Owen Weatherall, "Categories and the Foundations of Classical Field Theories", in E. Landry (ed.), *Categories for the Working Philosopher*, Oxford: Oxford University Press. (Forthcoming)

[144] Edward Witten, "Some Comments On String Dynamics", arXiv:hep-th/9507121, 1995.

[145] Edward Witten, "Reflections on the Fate of Spacetime", Physics Today, April 1996.

[146] Alberto Zaffaroni, "Introduction to the AdS/CFT Correspondence", Classical and Quantum Gravity 17, pp. 3571-3597, 2000.