The Dynamical Approach to Spin-2 Gravity

Kian Salimkhani*

Forthcoming in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* (post-peer-review, pre-copyedit version); final authenticated version is available online at: `https://doi.org/10.1016/j.shpsb.2020.05.002`

Abstract

This paper engages with the following closely related questions that have recently received some attention in the literature: (a) what is the status of the equivalence principle in general relativity (GR)?; (b) how does the metric field obtain its property of being able to act as a metric?; and (c) is the metric of GR derivative on the dynamics of the matter fields? The paper attempts to complement these debates by studying the spin-2 approach to (quantum) gravity. In particular, the paper argues that three lessons can be drawn from the spin-2 approach: (1) different from what is sometimes claimed in the literature, central aspects of the non-linear theory of GR are already derivable in classical spin-2 theory; in particular, 'universal coupling' can be considered a derived 'theorem' in both the classical and the quantum spin-2 approach; this provides new insights for the investigation of the equivalence principle; (2) the 'second miracle' that Read et al. argue characterises GR is explained in the classical as well as in the quantum version of the spin-2 approach; (3) the spin-2 approach allows for an ontological reduction of the metrical part of spacetime to the dynamics of matter fields.

1 Introduction

This paper engages with the following closely related questions that have recently received some attention in the literature on the foundations and interpretation of relativity theory (e.g., Read et al. (2018), Read (2018, 2019), Salimkhani (2018), and Lehmkuhl (2020)): (a) what is the status of the equivalence principle in general relativity (GR)?; (b) how does the metric field obtain its property of being able to act as a metric that is surveyed by material rods and clocks?; and (c) what is the ontological status of the metric in GR—i.e., is the metric ontologically independent and fundamental or is it derivative on the dynamics of the matter fields? I attempt to complement these debates by studying both the classical and the quantum version of the spin-2 approach to gravity. In particular, I argue that three lessons can be drawn from the spin-2 approach: (1) different from what is sometimes claimed in the literature, central aspects of the

*Philosophisches Seminar, Universität zu Köln, Albertus-Magnus-Platz 1, 50923 Cologne, Germany; k.salimkhani@uni-koeln.de.

non-linear theory of GR are already derivable in classical spin-2 theory; in particular, 'universal coupling' should not be considered a derived 'theorem' only in the context of relativistic quantum particle theory, but in the classical spin-2 approach as well; this provides new insights for the investigation of the equivalence principle; (2) the 'second miracle' that Read et al. (2018) argue characterises GR is explained in the classical as well as in the quantum version of the spin-2 approach; (3) the spin-2 approach allows for an ontological reduction of the metrical part of spacetime to the dynamics of matter fields; relatedly, chronogeometricity (i.e., the metric field's property of acting as a metric that is surveyed by material rods and clocks) receives a more rigorous (reductive) explanation.

The features of the classical and the quantum version of the spin-2 approach as well as their relations to Einstein's (non-linear) theory of general relativity have been worked out by a number of physicists, e.g., Markus Fierz and Wolfgang Pauli (1939), Nathan Rosen (1940*a*; 1940*b*), Achille Papapetrou (1948), Suraj Gupta (1952*b*; 1952*a*; 1954; 1957), Robert Kraichnan (1955; 1956), Walter Thirring (1959; 1961), Viktor Ogievetsky and Igor Polubarinov (1965), Walter Wyss (1965), J. Fang and Christian Frønsdal (1979), Richard Feynman (1995), Steven Weinberg (1964*a*; 1964*b*; 1972), Stanley Deser (1970; 2010), Robert Wald (1986), and Brian Pitts and William Schieve (2001*c*; 2007)—for a concise review see Preskill and Thorne (1995), for an extensive list of references see also Pitts (2016). In particular, it is suggested by Weinberg and others that the spin-2 approach to quantum gravity has a particularly significant feature: it is taken to provide an explanation for why the strong equivalence principle (SEP)—that is, the fact that locally all laws reduce to the laws of special relativity (SR)—holds (see Salimkhani (2018) for a sketch of the argument for a philosophical audience). I investigate these claims in sections 2 and 3. In short, Weinberg shows that a consistent, Lorentz-invariant quantum field theory (QFT) of massless spin-2 particles—i.e., gravitons—requires universal coupling of the graviton to all particles including gravitons. Thus, or so the reasoning goes, the SEP is to gravity what local gauge symmetries are to the other interactions: "quantum theories of mass zero, spin one particles violate Lorentz invariance unless the fields are coupled in a gauge invariant way, while quantum theories of mass zero, spin two particles violate Lorentz invariance unless the fields are coupled in a way that satisfies the equivalence principle" (Weinberg, 1999).

Proponents of this view argue that what makes this result significant is that it transforms a postulate of GR—i.e., something that needs to be put in by hand in addition to the dynamics—into a theorem deducible from the dynamics in QFT (e.g., Nicolis (2011)). I discuss this in more detail in sections 2 and 3 and also present some foundational and philosophical twists and consequences. In particular, I will argue that Weinberg does not straightforwardly provide a full-blown derivation of the standard version of the SEP in GR, and that what he does is—*pace* some comments in the literature[1]—in large parts already feasible in the classical spin-2 approach.

The latter is especially relevant with respect to recent work regarding the dynamical approach to GR which has its origins in the work of Harvey Brown and Oliver Pooley (see Brown and Pooley (2001, 2006), and Brown (2005)). In particular, it is proposed in this paper that the spin-2 results have the following ramifications for our understanding of gravitation in general

[1]For example, Carroll (2004) argues that one needs to "make the additional demand" (Brown (2005) speaks of an 'additional requirement') of self-coupling for the graviton, though in fact it is derived from the spin-2 theory being non-free and consistent.

and for GR specifically: (1) *pace* the standard 'geometrical' reading, it underlines the physicality (the 'field view') of the metric field in the spin-2 approach and in GR, which fuels an otherwise 'stuck'[2] dynamical understanding of the theory; and (2) it contributes to an understanding of (metrical aspects of) spacetime as non-fundamental. Let me briefly comment on these two claims in the following.

Canonically, in both SR and GR the metric field (plus the manifold) is identified with spacetime, giving rise to a geometrical understanding of the theories and—in the case of GR—a geometrical understanding of gravitation. Considering SR, for example, one can ask whether it is the Minkowskian nature of spacetime that explains "why the forces holding a rod together are Lorentz-invariant or the other way around" (Balashov and Janssen, 2003). For the standard geometrical approach—which (a) accepts the metric as part of the fundamental ontology and (b) argues that it has its chronogeometricity essentially—spacetime geometry "is the explanans here and the invariance of the forces the explanandum" (Balashov and Janssen, 2003).

Against this, the dynamical approach (see also section 4.1) argues that "it is wholly unclear how this geometrical explanation is supposed to work" (Brown, 2005, 134). Accordingly, the dynamical approach, in a sense, 'turns around' the explanatory relation between the symmetries of spacetime and the symmetries of the matter fields in the dynamics of the theory. This results in the following two central claims of the dynamical approach (see Brown and Read (2018), and Read (2019)): (a) the Minkowski metric $\eta_{\mu\nu}$ of SR can be ontologically reduced to the symmetry properties of the dynamics of matter fields (chronogeometricity then follows automatically); and (b) the (ontologically autonomous) metric field $g_{\mu\nu}$ of GR does not have its chronogeometricity essentially, but acquires it (from the SEP).[3]

There is a sense in which the issue of the metric's chronogeometricity is what the dynamical approach is after primarily, whereas the reduction story in SR is secondary; in this understanding, the failure to reduce the metric of GR is rendered irrelevant for the programme to stand. However, in trying to bring forward an explanation for the metric's chronogeometricity in GR, the proponent of the dynamical approach has to countenance two unexplained 'miracles', as Read et al. (2018) argue. In SR, on the contrary, only one miracle arises, precisely due to the ontological reduction. I shall therefore take the reduction claim as central for a full-fledged dynamical approach, and, hence, view the dynamical approach to GR as less satisfactory—i.e., as 'stuck'—compared to the dynamical approach to SR. A solution to this 'shortcoming' of the dynamical approach to GR is presented in section 4.2.

Focussing on the reduction aspect is in line with Norton's (2008) understanding of the dynamical approach as a 'constructivist' programme that is directed against the standard realist—or rather fundamentalist (see North (2018) and Menon (2019))—conception of spacetime (where spacetime structure, i.e., metric and manifold structure, is part of the fundamental ontology). While the geometrical view is the standard interpretation of GR in terms of *explanation*, spacetime fundamentalism may be taken as the standard interpretation of GR in terms of its *ontological*

[2]The dynamical approach to GR can be considered less successful than the dynamical approach to SR because it fails to ontologically reduce the GR metric g. Arguably, one does not have to take the reduction claim as the central claim of the programme (see below).

[3]The dynamical approach to relativity theory can be considered a philosophical variant of the 'field view' which is widely held among particle physicists. Against the geometrical understanding of GR, the field view takes the metric field in GR as—from the outset—a physical field like the others (e.g., Weinberg (1972)).

commitments. A spacetime constructivist—*pace* the fundamentalist—holds that spacetime (geometry) is fully reducible to the dynamics of matter fields. Indeed, at least within the dynamical approach to SR, spacetime is understood to arise from the dynamics of the matter fields. In other words, spatiotemporal structure is derivative on matter field dynamics.

It is here where the spin-2 approach to GR has similar ontological import. In particular, I argue that the dynamical interpretation suggested by the spin-2 approach contributes to an understanding of metrical aspects of GR spacetime as non-fundamental, but derivative on matter field dynamics. This resurrects the dynamical approach to GR in terms of some form of relationalism (see Pooley (2013), Sect. 6.3.2). This is especially significant as Brown (2005)—contrary to his stance on SR—renounces the ontological reduction aspect of the dynamical approach to GR and accepts the metric field g as ontologically autonomous, just as the geometrical approach has it.

The plan of the paper is as follows: In section 2, both the classical and the quantum spin-2 approaches are briefly presented. Then, in section 3, a first lesson is drawn regarding the status of the equivalence principle in these theories. In section 4, a recent take on the debate over the dynamical approach is discussed in light of the previous findings. In particular, I claim that the 'second miracle' made explicit in Read et al. (2018) disappears not only when moving to QFT, but even at the classical level. In section 5, I then argue for a third lesson, namely that the presented proposal can be cashed out in terms of an ontological reduction of spacetime. I conclude by discussing a few potential criticisms of my approach.

2 The Spin-2 Approach

2.1 The Classical Spin-2 Approach

It is sometimes implied that substantial insights from the spin-2 perspective are only to be expected in QFT (e.g., Nicolis (2011), Carlip (2014), Carroll (2004), and Salimkhani (2018)). It turns out, however, that most features of the spin-2 approach already appear in its classical version. In particular, the Einstein Field Equations (EFE) can be derived "non-geometrically" (Deser, 1970, 9)—i.e., without drawing on any geometrical notions, but by employing the standard field-theoretical techniques. This also gives a plausible explanation of why the particle physics perspective on GR should be so successful—it really is nothing but the quantised version of classical spin-2 gravity.

Usually, the 'classical spin-2 field'[4] is introduced in the linearised limit of Einstein's theory of

[4]Strictly speaking, spin is a quantum property. However, the spin properties of quantum fields express certain rotational symmetry properties that have a specific classical counterpart (see Maggiore (2008), Weinberg (1972; 1995) and Preskill and Thorne (1995)). Consider a generic classical second-rank tensor $X_{\mu\nu}$ with 16 components (or degrees of freedom). $X_{\mu\nu}$ decomposes into the following classical (i.e., non-unitary) irreducible representations of the Lorentz group: an antisymmetric tensor $A_{\mu\nu} = -A_{\nu\mu}$ (six independent degrees of freedom), a symmetric ($S_{\mu\nu} = S_{\nu\mu}$) and traceless ($S^{\mu}{}_{\mu} = 0$) tensor $S_{\mu\nu}$ (nine independent degrees of freedom), and a trace $X \equiv X^{\mu}{}_{\mu}$ (one degree of freedom). With respect to the rotation subgroup $SO(3)$, the antisymmetric tensor decomposes into two (generally massive) spin-1 vectors with three components ($2s + 1$ with $s = 1$); while the symmetric and traceless tensor decomposes into a spin-0 scalar, a spin-1 vector, and the five components ($2s + 1$ with $s = 2$) of a (generally massive) spin-2 tensor (Maggiore, 2008, 70). It is important to note that a massless spin-2 (which yields long-range interactions) only has two physically significant degrees of freedom. Accordingly, the symmetric

GR in terms of the symmetric and traceless tensor $h_{\mu\nu}$ representing a small perturbation on the Minkowski metric $\eta_{\mu\nu}$, i.e., an expansion of $g_{\mu\nu}$ to the first order in $h_{\mu\nu}$ (e.g., Maggiore (2008)):

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \ , \qquad \left|h_{\mu\nu}\right| \ll 1 \ . \tag{1}$$

This limit is convenient for certain analyses and applications, e.g., with respect to gravitational waves. However, classical spin-2 theory is not merely an *ex post* limit of GR, but a field theory in its own right that allows for substantial foundational insights. To see this, let us forget for a moment about GR, its weak field limit, the curved metric g, and the metrical interpretation of h, and consider a classical field theory of a free massless spin-2 field h on a fixed[5] flat Minkowski background η.[6] In this context, h is treated as any other field defined on flat Minkowski spacetime (e.g., the electromagnetic field). In particular, as in any other field theory, h is assumed to vanish at infinity,[7] but *not* assumed to be small (see Deser (1970; 1987; 2010); see also Weinberg (1972, 165)). In particular, we shall find that this leads us to the full[8] non-linear EFE.

Historically, several approaches—different in strength—have been pursued (see Preskill and Thorne (1995), and Fang and Frønsdal (1979) for an overview). In particular, Gupta (1954) first proposed that one may derive the non-linear EFE from a linear massless spin-2 theory (building on Fierz and Pauli (1939)) by iteratively summing an infinite series of energy-momentum contributions (see below; see also Fang and Frønsdal (1979)). It is important to note, however, that Gupta—like others who worked out similar accounts (e.g., Feynman)—did not carry out the calculation in full, and, hence, did not properly *prove* that Einstein's theory can be obtained in this way; it is Kraichnan (1955) who first (and independently) obtained the EFE. Notably, it is Deser (1970) who gave the first rigorous and complete derivation arriving at the full theory in closed form (i.e., in one step without iteration). In brief, Deser bypasses the infinite series summation (only an additional cubic term appears) by using the Palatini formalism, i.e., by treating the metric and the affine connection independently. Deser's (1970) derivation is also considered to be the stronger result with respect to presuppositions and scope: in particular, unlike Gupta (1954), Deser (as well as Kraichnan and Feynman, for example) does not simply *assume* that gravity couples to the total energy momentum tensor, but *derives* this from the consistency of the field equations (Preskill and Thorne, 1995).[9]

---

and traceless tensor represents a massless spin-2 field for appropriate invariance conditions (i.e., gauge invariance) that eliminate the seven spurious degrees of freedom (see also Ogievetsky and Polubarinov (1965)). In this sense, we can call the (classical) symmetric and traceless tensor field $S_{\mu\nu}$ corresponding to the massless spin-2 particle a 'classical spin-2 field'. It is usually denoted as $h_{\mu\nu}$. Note that for technical reasons h is often taken to represent a symmetric but not traceless tensor with eight degrees of freedom to be eliminated (see Maggiore (2008, 71) and Fang and Frønsdal (1979, 2264–2265)).

[5]Here, $\eta_{\mu\nu}$ is a fixed field in the sense of Pooley (2017, 115): it is identically the same in all kinematically possible models.

[6]For reviews see, for example, Misner et al. (1973), Alvarez (1989), Preskill and Thorne (1995), Sexl and Urbantke (2002), Maggiore (2008), and Ortín (2015); original works are due, for example, to Gupta (1954), Kraichnan (1955), Deser (1970; 2010), Thirring (1959; 1961), Ogievetsky and Polubarinov (1965), Wyss (1965), Fang and Frønsdal (1979), Wald (1986), and Pitts and Schieve (2001*c*).

[7]In a flat background setting, this standard field-theoretic assumption prevents the derivation of the cosmological constant term (Deser, 1987, L103); see also below.

[8]A few qualifications will be briefly discussed at the end of this section.

[9]Furthermore, Deser (1970) does not need to introduce a particular gauge. See also Deser's (2010) reply to Padmanabhan (2008).

While the attempts at deriving the EFE in the spin-2 approach differ with respect to technical details, presuppositions, rigour, and completeness, the reasoning is essentially the same: the full non-linear theory (i.e., GR) is argued to follow from the linear free field equations for a massless, spin-2 field h since these equations pose non-trivial consistency conditions that are only[10] satisfied by the full theory. To be a bit more specific, the derivation of the EFE can either focus on gauge invariance (e.g., Thirring (1959; 1961), Ogievetsky and Polubarinov (1965), and Sexl and Urbantke (2002)), or on the mathematical self-consistency of the equations (e.g., Deser (1970); see Ortín (2015, 89)), but ultimately both gauge invariance and self-consistency are part of a full derivation (gauge invariance specifies the linear free field equations to which self-consistency is applied).[11] In this paper I focus on the self-consistency argument which, by exhibiting the non-linearities in terms of self-interactions of the (linear, i.e., formerly free) h field (Deser, 1970, 9), is more easily accessible. So as to pay particular attention to the relevant aspects for my discussion, I will not give a full derivation of the EFE here, but only sketch the general rationale.

The linear free field equations for a massless, spin-2 field h, are as follows:[12]

$$\left[(\eta_\mu{}^\alpha \eta_\nu{}^\beta - \eta_{\mu\nu}\eta^{\alpha\beta})\Box + \eta_{\mu\nu}\partial^\alpha\partial^\beta + \eta^{\alpha\beta}\partial_\mu\partial_\nu - \eta_\mu{}^\beta\partial^\alpha\partial_\nu - \eta_\nu{}^\alpha\partial^\beta\partial_\mu\right] h_{\alpha\beta} \quad = \quad 0 \, , \tag{2}$$

where $\Box = \partial_t^2 - \partial^i\partial_i$ is the d'Alembert or 'box' operator.[13] As mentioned, h is supposed to be a *free* field on η. Accordingly, the equation for h is consistent as it stands; its solutions are the solutions to the wave equation.

Now, we want to go beyond the free theory and introduce interactions between h and other (ordinary) matter fields. Technically, this is done by adding a dynamical source term on the right-hand side of the equation, the (ordinary) matter energy-momentum tensor $T_{\mu\nu}$. Note, however, that the left-hand side of the free field equation is divergenceless. This 'divergence identity' of the linear free field equations—which can be identified with the linearised Bianchi identity (Wald, 1986)—poses a severe constraint on any generalisation of the linear free field equations: any modification of the right-hand side must respect this constraint. In particular,

[10] All attempts at deriving the EFE demonstrate that *only* by including all contributions (the 'full' theory) are the consistency conditions satisfied. Some (e.g., Ogievetsky and Polubarinov (1965), Deser (1970), and Fang and Frønsdal (1979), but not Gupta (1954)) even argue that this *uniquely* leads to GR—meaning that "Einstein's nonlinear theory of gravitation is the only consistent, Lorentz-invariant theory of an interacting, massless, spin-2 field in flat space" (Fang and Frønsdal, 1979, 2267f). This stronger claim might even be considered the received view today (see Preskill and Thorne (1995)). However, such claims need to be evaluated carefully as, for example, Ortín (2015; 2017) argues. In particular, some additional presumptions will typically be made (e.g., Ogievetsky and Polubarinov (1965)). In this paper, I will not evaluate the uniqueness claims.

[11] I thank Brian Pitts for this.

[12] See Alvarez (1989) for more details.

[13] Deser (1970) does not provide an argument on how to obtain these equations, but seems to rely on the linear limit of the vacuum field equations, i.e., $G^{(L)}_{\mu\nu}(h) \equiv R^{(L)}_{\mu\nu}(h) - \frac{1}{2}R^{(L)}(h)\eta_{\mu\nu} = 0$, where $G^{(L)}_{\mu\nu}(h)$, $R^{(L)}_{\mu\nu}(h)$, and $R^{(L)}(h)$ are the linearised Einstein tensor, the linearised Ricci tensor, and the linearised Ricci scalar for h, respectively. With respect to the philosophical issues this paper deals with, it is important to note that this does not pose any circularity problem: the dependence of the linear free field equations for h on the full non-linear vacuum field equations for g is merely heuristical. In fact, the linear free field equations for h can be obtained independently from GR by arguing for the simplest linear spin-2 theory in analogy with electromagnetism (see Fierz and Pauli (1939), Feynman et al. (1995), Alvarez (1989), and Ortín (2015)). Hence, at first sight, it seems plausible that it is the derivation of the full dynamical field equations from the linear free field equations that determines the ontological relationship between h and g (cf. section 6).

a dynamical source in terms of an energy-momentum tensor must have vanishing divergence, i.e., must be conserved.[14] Adding $T_{\mu\nu}$ as a dynamical source on the right-hand side, however, introduces a term with non-vanishing divergence: coupling h to ordinary matter implies—as is derivable from the equations of motion—that the ordinary matter energy-momentum tensor is no longer conserved ($T^{\mu\nu}{}_{,\nu} = \nabla_\nu T^{\mu\nu} \neq 0$).[15] 'Naively' coupling h to other matter fields by adding $T_{\mu\nu}$ on the right-hand side renders the equations inconsistent. By coupling h to ordinary matter, energy and momentum can be exchanged between h and the ordinary matter fields, thus, energy-momentum conservation no longer holds for the ordinary matter fields alone.

To do away with this inconsistency, conservation of energy and momentum needs to be restored by including the self-coupling contributions of h to the *total* energy-momentum tensor (i.e., by adding them to the ordinary matter energy-momentum tensor $T_{\mu\nu}$). For the linear theory introduced above (obtained from a Lagrangian quadratic in h), this means that we need to include the quadratic contribution ${}^{(2)}\Theta_{\mu\nu}(h)$. However, now the formerly linear equations are rendered quadratic, which is why we actually need to consider the quadratic equations of motion, i.e., the Lagrangian cubic in h, which contributes the cubic energy-momentum tensor ${}^{(3)}\Theta_{\mu\nu}(h)$, which renders the equations of motion cubic, i.e., the Lagrangian quartic in h, and so on. In this way, an infinite series of contributions to the energy-momentum tensor is forced on us by a mathematical consistency condition (the divergence identity of the linear free field equations for h). Starting from the linear theory, it turns out not to be sufficient simply to include the self-coupling contributions of that linear theory to restore consistency; only for the full series of self-coupling contributions ($\Theta_{\mu\nu} = \sum_{n=2}^{\infty} {}^{(n)}\Theta_{\mu\nu}$) is energy-momentum conserved, and, hence, the consistency condition satisfied.

In this way, we obtain an infinite series of self-coupling contributions of h which can be shown to sum to the full non-linear Einstein equations $G_{\mu\nu}(\eta + h) = -\kappa T_{\mu\nu}$ (Deser, 1970). The left-hand side only contains terms depending on $\eta_{\mu\nu} + h_{\mu\nu} \equiv g_{\mu\nu}$, which can be reexpressed accordingly. All non-linear contributions ($\mathcal{O}(h^2)$ and higher), i.e., all self-coupling contributions of h, are absorbed in the now-conserved $T_{\mu\nu}$ on the right-hand side.[16] Note again that $\eta + h$ is *not* a first-order expansion of g; h is *not* a small perturbation on η but a field in its own right that is then interpreted as the *full* deviation of a (new) metric g from η (see Deser (1970; 2010)).

To summarise, if *any* matter coupling of h is to be allowed, the self-coupling contributions of h need to be included for consistency reasons. Either the spin-2 theory is trivialised (stays free), or the h field is required to couple to *all* fields present (including itself), i.e., the *total* energy-momentum tensor of the theory, with the same strength (Ortín, 2015, 78). Any partial or non-universal coupling to ordinary matter would render the theory inconsistent (see also Kraichnan (1956)). Establishing that h couples universally to all energy and momentum is then interpreted as establishing the strong equivalence principle (see section 3) and as suggesting—*ex post*—a geometrical interpretation of the (from the outset *non*-geometrical) theory:

> Consistency has therefore led us to universal coupling, which implies the equivalence

[14] It is interesting to note that a scalar field theory is significantly less restrictive. In particular, it does not constrain which energy-momentum tensor is needed (Ortín, 2015, 47).

[15] Indeed with respect to the ordinary (not the covariant) divergence (Weinberg, 1965*b*, 471).

[16] To follow the rather brief exposition of the argument in the paper at hand, as well as in the works cited, it is instructive to consider Weinberg's (1972, 165f) comments on how to separate linear from non-linear contributions of h in the Ricci tensor.

principle. It is at this point that the geometrical interpretation of general relativity arises, since all matter now moves in an effective Riemann space of metric $g^{\mu\nu} \equiv \eta^{\mu\nu} + h^{\mu\nu}$, and so the initial flat 'background' space $\eta^{\mu\nu}$ is no longer observable. (Deser, 1970, 13)

I investigate this aspect further in section 5.

To conclude, the most important philosophical insight is that the following two theoretical set-ups turn out to be equivalent (up to a few qualifications discussed below)—both feature the metric g (either built-in or effectively) and both yield the non-linear EFE. In the first (standard) set-up, one has (apart from the ordinary matter fields and their dynamics): (a) the (generally curved) g field, and (b) the presumption of the SEP; in a second (alternative) set-up, one has (again, apart from the ordinary matter fields and their dynamics): (a) the (flat) η field, and (b) the h field which couples universally to all energy and momentum including self-coupling.

Note that in the appropriate dynamical reading the *fundamental* assumptions (i.e., the Quinean ontology and ideology) of these two set-ups are, in case (1), the g field and the SEP, and, in case (2), the additional matter field h and its Lorentz-invariant dynamics. (In a dynamical reading, the fixed, flat η field is ontologically reduced to properties of the matter fields (see section 4).)

Looking ahead, the classical spin-2 derivation of full non-linear GR (up to a few qualifications to be discussed in a moment), including a derivation of the universal coupling of h, will come in handy to shed new light on a couple of foundational and philosophical issues in GR. In particular, as I argue further in section 5, an ontological reduction of g is suggested. In light of this feature of the classical spin-2 approach, neither the SEP, nor the g field, nor GR as a theory are fundamental, but are rather derivable from a special-relativistic spin-2 field theory.

Before I briefly review the quantum version of the spin-2 approach, let me point out a few qualifications regarding the presentation above.

First, the flat background derivation presented (using a flat metric η defined on a flat manifold) is *not* able to accommodate the cosmological constant term $\Lambda g_{\mu\nu}$ of the full EFE. This is at odds with the claim that flat spin-2 theory is equivalent to full GR.[17] Technically, the issue is that $\sqrt{g}R^{\mu\nu}(g) = \lambda g^{\mu\nu}$ needs to be fulfilled. However, "the only way to get $\sqrt{g}$ from flat space is to put in by hand a source $\eta_{\mu\nu}$ of the initial linear equations", which "violates the assumed fall-off at infinity of the h field" (Deser, 1987, L103). To be able to construct a cosmological term in the spin-2 framework, we need to start out from a more general—i.e., curved—background space, namely from an *Einstein space*[18] (with a Lorentzian signature). In fact, Deser (1987) and Deser and Henneaux (2007) show that the derivation presented for a flat (i.e., $R^{\mu}{}_{\nu\sigma\rho} = 0$ everywhere) background η can be generalised to the class of curved backgrounds $\tilde{g}$ with $R_{\mu\nu} \propto \tilde{g}_{\mu\nu}$. This

[17] This might be considered a minor issue, though, since one may very well argue that setting the cosmological term to zero is justified because Λ is empirically constrained to be very small (Weinberg, 1972).

[18] An *Einstein space* is a generally curved (pseudo-)Riemannian differentiable manifold with the property that its Ricci tensor is proportional to its metric, i.e., $R_{\mu\nu} = \lambda g_{\mu\nu}$ for some constant λ (Deser, 1987). Hence, for an Einstein space the (generally curved) metric solves the full vacuum EFE (including a cosmological constant term which effectively acts as a source such that the energy-momentum tensor is proportional to the metric). Einstein manifolds with $\lambda = 0$ are called *Ricci-flat*. Note that a Ricci-flat manifold (i.e., $R_{\mu\nu} = 0$) does not need to be flat with respect to curvature (i.e., generally $R^{\mu}{}_{\nu\sigma\rho} \neq 0$). The flat manifold with its flat metric η used throughout this section is flat in the strictest sense: all components of the curvature tensor vanish everywhere.

permits the construction of the cosmological term (Deser, 1987).

Second, given these assumptions with respect to the background manifolds we start from, one might worry whether certain topological or causal restrictions with respect to the solution space may enter: although, formally, the full GR action is obtained, the solution space—i.e., the space of recoverable g fields—might be different from full GR due to such restrictions being inherited. However, it seems that no severe topological or causal restrictions do apply. At most, some moderate constraints for h might enter when one (rightfully) demands the retention of the correct signature of the effective metric g, or the correct relation of the two respective null cones (such that the effective g-causality does not violate the background η-causality; see Penrose (1980), Pitts and Schieve (2001*a*), and Pitts and Schieve (2001*b*)). The latter arguably results in a restriction to globally hyperbolic solutions.[19]

Third, and more as a clarifying remark, one might worry similarly that deriving the EFE perturbatively restricts the solution space to analytic solutions (i.e., solutions expressible by a convergent power series). It may be that some approaches, like Gupta's (1954), are affected by this, however, most classical derivations are not perturbative in the sense of adding orders of a small perturbation; in particular, Deser's closed form derivation is not (and neither is Kraichnan's (1955) derivation). Accordingly, such restrictions do not apply. For quantum derivations the situation will generally be different (see end of next section).

2.2 The Spin-2 Approach to Quantum Gravity

Regarding the general rationale, the different quantum derivations proceed essentially in the same manner: the consistency of some general set-up for an interacting spin-2 particle—usually referred to as the 'graviton'—requires specific coupling properties. Similar to the classical case, universal coupling of the graviton to *all* particles that have energy and momentum is established as the result of postulating consistent Lorentz-invariant graviton interactions with *some* matter. For the details of the (different) original derivations see Weinberg (1964*a*; 1964*b*; 1965*a*; 1965*b*), Boulware and Deser (1975), and Davies and Fang (1982). In the following, I shall only briefly review how universal coupling is obtained in Weinberg's (1995) 'soft graviton argument',[20] as I shall then be more concerned with the foundational and philosophical consequences of this result.

Note, first, that in principle one could start from the previously derived classical theory and attempt to quantise it as a non-abelian gauge theory. One will then soon run into the well-known problems of the theory's non-renormalisability. For high energies, perturbative quantum gravity is non-predictive and requires modifications. In the low-energy regime, however, the quantised theory perfectly agrees with the results of classical GR and all observational data.

However, there is also

> an entirely different possible approach to quantum gravitation which avoids the high frequency problems of local field theory; it begins with the firmly established principles of special relativistic quantum particle (rather than field) theory, namely Lorentz

[19] I thank Brian Pitts for helpful discussions.

[20] Here, 'soft' refers to the gravitons being low-energetic. For accessible reassessments of this argument, see Nicolis (2011) or Salimkhani (2018).

invariance and the postulate that all forces are transmitted by the virtual exchange of particles. (Boulware and Deser, 1975, 194)

Weinberg's soft graviton argument is of this kind. Similarly, Boulware and Deser (1975) derive GR from general properties of special-relativistic particle interactions.

Weinberg considers a generic Lorentz-invariant scattering process $\alpha \to \beta$ with some non-vanishing transition amplitude $\mathcal{M}_{\alpha\beta}$ and modifies it to include the emission of a soft massless spin-2 particle with momentum $q \to 0$ from one of the in- or outgoing states, i.e., some arbitrary particle i with momentum p_i^μ. Now, such a modification ought not to spoil Lorentz invariance. Hence, the emission—more specifically, the leading order contribution of the spin-2 emission—needs to preserve Lorentz invariance independently. This then yields that the sum over all particle momenta p_i^μ weighted by their couplings to the spin-2 particle κ_i vanishes, i.e., $\sum_i \kappa_i p_i^\mu = 0$.

However, (unweighted) momentum conservation, $\sum_i p_i = 0$, already holds. Hence, either the scattering process is trivial (the particles do not interact, i.e., the individual momenta p_i^μ do not change at all), or the coupling constants κ_i are equal for all particle species regardless of their properties, i.e., $\kappa_i = \kappa$.

To summarise, the soft graviton argument shows that Lorentz invariance forces massless spin-2 particles to universally couple to other particles. Note that no restriction applies to the notion of 'other particles' (nor to 'universal'). Hence, the established 'universal coupling' implies graviton self-coupling. In section 3, I investigate claims to the effect that universal coupling implies the SEP.

The derivations of universal coupling by Weinberg or Boulware and Deser have several other important features. For example, it directly follows that the massless spin-2 particle is unique: if without any assumption about the particle species the coupling constant is universal, 'another' massless spin-2 field would have the exact same properties and, therefore, be indistinguishable from the graviton. Two universal interactions cannot be told apart. It also follows that gravitation cannot be a higher spin effect—i.e., there are no higher spin (tensor) contributions to the long-range behaviour of gravitation (see Weinberg (1965*a*; 1965*b*; 1995) and Boulware and Deser (1975)): otherwise this would put another (weighted) constraint on the summation of the momenta, which could not be met unless the corresponding couplings universally vanish. Moreover, there are no conserved symmetric tensors of rank $r \geq 3$ that could act as a source—except for the total derivative of an asymmetric tensor (see Weinberg (1965*a*; 1965*b*; 1995)). However, these so-called 'Pauli-type' tensors (due to Pauli (1941)), which are sometimes referred to as non-minimal coupling terms (e.g., Boulware and Deser (1975)), yield amplitudes that vanish in the infrared limit and only contribute to the high-energy behaviour. Accordingly, particles of spin $j \geq 3$ can exist, but cannot mediate inverse square law forces or generally have couplings that are present at low-energies.[21]

Before we move on, recall that universal coupling including self-coupling is already derived in the classical spin-2 approach. So, in a sense, the (independent) special-relativistic quantum particle derivation merely confirms that the quantised theory also has this feature.[22] However,

[21]Another way to put this is that all these modifications will behave as high-energy corrections to the GR limit.

[22]With respect to unification, it is certainly a philosophically interesting observation that GR can be constructed from QFT independently (see Salimkhani (2018)).

the quantum version may still be considered more important. This is because quantum theory is generally assumed to be more fundamental than GR. Hence, one may argue that the (more fundamental) quantum result is, ultimately, *the reason* for the result obtaining in the (higher-level) classical spin-2 theory. In particular, one may argue that it is the quantum result that selects the preferred ontology, not standard GR or the classical spin-2 view (which on its own is on a par with the standard GR view).

Furthermore, the relativistic quantum framework seems also generally more restrictive than classical field theory, for example, due to the results of Weinberg's (1964*b*) low-energy theorem and the Coleman-Mandula (1967) no-go theorem based on a few general assumptions.[23] If this is correct, the real constraint is the special-relativistic quantum particle derivation[24]—which, as mentioned, naturally drops out of QFT, but is not tied to QFT since it is more generic.[25]

Now, as soon as universal coupling is established, the story is basically the same as in the classical case. In particular, the full non-linear classical theory can be constructed from here (a few qualifications will be discussed in a moment). Accordingly, all further philosophical observations also apply. In particular, given the more fundamental framework, the quantum spin-2 theory can be viewed as decisively supporting an ontological reduction of g. Also, since GR can be obtained independently as the classical limit, GR is arguably reduced to QFT (see Salimkhani (2018)) and may be viewed as non-fundamental, i.e., as

> an essentially phenomenological (albeit unique) theory for describing interactions at macroscopic distances and times. (Boulware and Deser, 1975, 230)

It is particularly the quantum spin-2 approach that underlines and makes precise the effective field theory viewpoint of GR (see Donoghue (1994)). Hence, Lehmkuhl's (2019) classification of GR as a hybrid theory is called into question: GR is fundamental neither regarding matter (which is uncontroversial), nor regarding gravity.[26]

Finally, a few qualifications need to be discussed. First and foremost, the effective field theory concept already implies that the quantum spin-2 approach is to be understood as yielding the low-energy limit of some full-fledged theory of quantum gravity. More precisely, Boulware and Deser (1975) argue that the special-relativistic quantum particle derivation is "in no way restricted to 'weak field' situations ... but only problems involving slow variation compared to

[23] Regarding what survives long-range, the relativistic quantum framework only allows for massless fields with spins $0, 1/2, 1, 3/2$, and 2.

[24] Weinberg seems to hold this view: "In my opinion the answer is not to be found in the realm of classical physics, and certainly not in Riemannian geometry, but in the constraints imposed by the quantum theory of gravitation" (Weinberg, 1972).

[25] Note also that the spin-2 particle is necessarily part of string theory, which arguably further bolsters the spin-2 view.

[26] As is well known, one can distinguish between 'principle' and 'constructive' theories. A constructive theory is usually understood to have higher explanatory power than a principle theory (see Brown and Pooley (2006)), as it draws on fundamental structures only. Similarly to SR, GR can be viewed as a principle theory, at least in part, although this is controversial. Lehmkuhl (2019) dubs it a 'hybrid theory': fundamental regarding gravity, effective and phenomenological (and hence—or so I shall take it—more like a principle theory) regarding matter. There is a sense in which the theory is based on a number of postulated principles—that is, according to Einstein (1914), precisely formulated empirical facts. In particular, the key principle for the strong connection between gravity and spacetime in standard GR is the equivalence principle (e.g., Poisson and Will (2014)).

particle Compton wavelengths."[27] Similarly, Weinberg's (1965*a*) derivation of the EFE is not restricted to the soft particle case, but relies on the more robust Dyson-Feynman perturbation theory.[28]

Second, a (so far neglected) cosmological term can be included by means of a specific matter-like contribution to the energy-momentum tensor (Boulware and Deser, 1975, 197).

Third, one might consider it a shortcoming of Weinberg's argument that it depends on a specific class of backgrounds. On the one hand, the geometry needs to be flat and Minkowskian (such that Lorentz invariance can be employed straightforwardly), and on the other, the topology needs to allow for the construction of an S-matrix (which excludes topologies without an asymptotic regime) (Davies and Fang, 1982, 472). However, as Davies and Fang (1982) show, Weinberg's derivation can be generalised by directly working with the action.[29]

Fourth, all mentioned quantum derivations are manifestly perturbative. As such, they rely on certain analyticity assumptions (Boulware and Deser, 1975, 194). As mentioned before, this will typically restrict the solution space to analytic solutions. Other classes of solutions are, *prima facie*, not mathematically rigorously recoverable. So, within such derivations one would need to have some independent argument for why the recovered solution space may, nevertheless, be extrapolated to the full solution space of GR.

To add a last remark, having a restricted solution space is generally not considered to render a theory physically uninteresting *per se*. Quite the opposite, in case the restrictions only exclude solutions that are taken to be unphysical anyway—which would need to be investigated.

3 Lesson I: A Closer Look at the Equivalence Principle

When discussing the two variants of the spin-2 approach I did not further specify what the spin-2 field's 'universal coupling' amounts to with respect to the equivalence principle of GR. In the physics literature, it is frequently claimed—sometimes explicitly, sometimes implicitly—that the spin-2's universal coupling *is* (variously) the 'equivalence principle', the 'weak equivalence principle', and the 'strong equivalence principle'. This ambiguity cries out for some further philosophical clarification—especially as the equivalence principle is in many ways the key to interpreting GR.

Let me first recall what the equivalence principle in GR is standardly taken to mean, and distinguish the following traditional versions:[30]

[27] In fact, due to this generality, the neglect of higher-order interactions therefore requires a further justification, which they do also provide. In short, higher-order terms are negligible as long as quantum corrections to the lower-order vertices are as well (Boulware and Deser, 1975, 197).

[28] I thank Nick Huggett for reminding me of this.

[29] Again, the starting point is the free field structure which "imposes a remarkably restrictive set of conditions on the structure of the higher-order terms" (Davies and Fang, 1982, 473). To lowest order, varying the full Lagrangian ($\delta L/\delta h_{\mu\nu} = 0$) yields the (free) Fierz-Pauli equation which, via the familiar divergence identity (see section 2.1), forces the higher orders to be divergenceless as well, order by order (Davies and Fang, 1982).

[30] Note that all the principles of equivalence come in various versions which I will not discuss here (for reviews of the systematic, experimental, and historical details, see, for example, Ghins and Budden (2001), Di Casola et al. (2015), Will (2001), and Lehmkuhl (2020)).

WEP-GR The free-fall trajectories of test bodies are independent of the test-body's composition.

EEP-GR All non-gravitational laws of physics are the same in every freely falling frame.

SEP-GR Locally, all laws of physics reduce to the laws of SR.

For this paper, I focus on the weak equivalence principle of GR (WEP-GR) and the strong equivalence principle of GR (SEP-GR).[31] According to WEP-GR—which is enshrined in the geodesic principle[32] (see Di Casola et al. (2015))—it is impossible to decide whether the observed effects on freely falling test particles stem from a gravitational field or from being situated in a uniformly accelerated frame. SEP-GR is usually understood to generalise this to a stronger constraint. It implies that any gravitational field can be 'transformed away' locally,[33] and is often taken to imply a reduction of gravitation to spacetime geometry. In fact, SEP-GR and many alternative formulations of it employ geometrical language or suggest a geometrical reading (for obvious reasons—after all, this is the standard formulation and interpretation of GR).

So, on a general level, 'WEP-like' principles refer to some form of composition-independence of a body's response to gravitation, and 'SEP-like' principles refer to the fact that laws of nature are locally special-relativistic. Since 'composition' (and 'locality') can be spelled out differently, one obtains several versions of WEP-like (and SEP-like) principles of different strength. Note that, at first sight, universal coupling of the spin-2 field (or particle) seems to resonate well with WEP-like principles that refer to concrete physical objects, but less so with SEP-like principles that refer to locality conditions on laws.

But before we relate the concepts of GR to the concepts of the spin-2 approach, it is useful to further distinguish variants of SEP-GR that are analysed with respect to other notions. First, SEP-GR has been analysed in terms of *Lorentz invariance*—or more precisely: Poincaré invariance (see Read et al. (2018)). Second, SEP-GR has been analysed as Lorentz invariance plus *minimal coupling*—i.e., no coupling of matter fields to the curvature tensor or its contractions (see Brown (2005)). Third, SEP-GR has also been analysed in terms of *universal coupling* to g—i.e., "[t]he idea that a single field g interacts with all the nongravitational[34] fields in a unique manner" (Will, 2001, 17)—plus minimal coupling (e.g., Will (2001) and Brown (2005)). It is important to note that these three variants are not equivalent, as Read et al. (2018) show. Minimal coupling is an independent condition that can actually lead to violations of SEP-GR (see Read et al. (2018)). This is why Read et al. (2018) drop minimal coupling in favour of Lorentz invariance as the sole condition.

Now, when turning to QFT or non-geometrical theories in general we would expect the geometrical coating to come off. It is here where reformulations of the equivalence principles in terms of concepts like Lorentz invariance, universal coupling (to g), and minimal coupling are particularly useful. Indeed, to capture the general physical content of the equivalence principles

[31] What is often dubbed the Einstein equivalence principle (EEP-GR) leaves open whether the (non-gravitational) laws of physics are special-relativistic or not.

[32] The geodesic principle of GR states that "free massive point particles traverse timelike geodesics" and, hence, can be thought of as a relativistic version of Newton's first law of motion (Malament, 2012, 245).

[33] Note that the notion 'local' may be cashed out in a 'pointy' or a 'neighbourhood' version (see Read et al. (2018)).

[34] We shall soon encounter another version of universal coupling without this restriction.

independently of geometrical terms we usually allude to gravity's universality: *all* objects are subject to gravity in the same way.[35] It is this property of gravitation that suggested geometrisation in the first place (in terms of it being an explanation): gravitational effects being apparently independent of the objects' properties supports the inference that gravitation arises from something else, namely spacetime.[36] But it also resonates well with the framework of non-geometrical theories like QFT, where we can account for gravity's universality in terms of universal coupling (to the spin-2 graviton).

Given the results of section 2, I distinguish the following versions of 'universal coupling' for the contexts of the classical spin-2 theory (UC-C), and the quantum spin-2 theory (UC-Q), respectively:

UC-C To obtain Lorentz-invariant interactions, the classical spin-2 field h defined on a fixed Minkowski background η is required to couple universally to the total energy-momentum tensor.

UC-Q To obtain Lorentz-invariant interactions, gravitons on a fixed Minkowski background η are required to couple universally to the total energy-momentum tensor.

What, then, is the relation between these universal coupling principles for the spin-2 theories and the traditional versions of the equivalence principles in standard GR? Can we show that the universal coupling principles for the spin-2 theories imply the traditional standard GR principles (or one specific traditional standard GR principle)? In particular, can we show that the universal coupling principles for the spin-2 theories imply the strongest GR principle, i.e., SEP-GR? Or can we at least argue that the fact that some particular GR equivalence principle holds is explained by the fact that some version of universal coupling for the spin-2 theories holds? As mentioned, the literature is quick to identify universal coupling (to the spin-2) with SEP-GR. Is this identification warranted?

A natural idea would be to first relate universal coupling of the (quantum) spin-2 field to universal coupling of g, and then argue that SEP-GR follows. Note, however, that universal coupling of g might indeed be viewed simply as one of the requirements for SEP-GR, but there may be other requirements as well (similarly, see Ortín (2017)). For example, one might want to additionally exclude certain types of (non-dynamical) fixed fields from the theory (e.g., fixed timelike vector fields), as they can violate Lorentz invariance.[37] Indeed, as mentioned, SEP-GR has been analysed as 'universal plus minimal coupling' (to g) by Brown (2005). However, Read et al. (2018) show that this is not correct. While universal coupling (to g) is necessary but not sufficient for SEP-GR, minimal coupling (to g) is a stronger constraint that, as Read et al. (2018) show, is neither sufficient nor necessary, but may violate certain versions of SEP-GR.

[35] This is what is behind some non-standard formulations of the SEP in terms of gravitational interaction, for example: "In general relativity, the principle of equivalence states that it is the total energy that interacts gravitationally and not just the bare mass" (Arnowitt et al., 2004).

[36] "The property that all non-gravitational fields should couple in the same manner to a single gravitational field—universal coupling—allows one [to] discuss the metric g as a property of spacetime itself rather than as a field over spacetime. This is because its properties may be measured and studied using a variety of different experimental devices, composed of different non-gravitational fields and particles, and because of universal coupling, the results will be independent of the device." (Will, 2001, 61)

[37] I thank an anonymous referee for this.

In Salimkhani (2018) I have previously admitted that it is not obvious which version of the equivalence principles Weinberg obtains, and have argued that Weinberg's UC-Q is best identified with WEP-GR. This does seem in line with Weinberg's (1964*a*) own understanding. But if that is correct, we face the problem that WEP-GR is usually understood to not suffice for constructing GR. So how do we obtain SEP-GR in the QFT setting—at least approximately? According to my earlier work, SEP-GR may be taken to follow since the often-invoked additional requirement of minimal coupling "is fulfilled here because all terms violating SEP [SEP-GR] essentially behave as high-energy corrections and are therefore absent in the low-energy limit. In this sense, universal coupling [of the graviton, identified with WEP-SR,] effectively implies the SEP [SEP-GR]" (Salimkhani, 2018, 35). This is similar to Schiff's conjecture that WEP-GR implies the stronger versions (see Schiff (1960); see also Read (2018)).[38]

The problem with this is that (1) Salimkhani (2018) does not really provide an argument for why UC-Q should be identified with WEP-GR, (2) minimal coupling is actually neither necessary nor sufficient for SEP-GR (Read et al., 2018), and (3) the different contexts seem to be slightly mixed up.

Instead, consider the following argument (elaborated further in section 5). Universal coupling of the spin-2 field (or particle) along with the fact that all physical fields are defined on η implies the universal coupling principle for g in GR, because g inherits the universal coupling property from h.[39] Furthermore, note that in both versions of the spin-2 theory minimal coupling (regarding η) is manifest for all fields (including the spin-2 field) *because they are from the outset nothing but special-relativistic theories.* Minimal coupling in the spin-2 context (i.e., with respect to η) is not an additional presupposition, because it is included in the assumption of (global) Lorentz invariance.[40] Note that this is not the condition revoked by Read et al. (2018) which refers to the GR context.

So, UC-C and UC-Q imply universal coupling of g in GR, which can indeed be identified with a strong version of WEP-GR; but what about SEP-GR?

Here, a few important caveats need to be addressed. An immediate worry is the following: why does the assumption of (global) Lorentz-invariant interactions in the understanding described above not already trivially imply SEP-GR (i.e., locally all laws reduce to the laws of SR)? For in a special-relativistic set-up, global Lorentz invariance *strictly implies* local Lorentz invariance. Then—differently to what is sometimes claimed in the physics literature (e.g., Nicolis (2011))—SEP-GR would have been a presupposition rather than a result.[41][42] Also (depending on the particular logical relations between the different principles under consideration), other versions of the equivalence principle—or indeed universal coupling—may *not* be viewed as being derived in any substantial sense either. If the SEP is presumed and universal coupling is (strictly)

[38] Note that in this context we are able to clarify the plausibility argument Schiff provides.

[39] The assertion, yet to be established, that a full-blown dynamical approach to spin-2 gravity is feasible, further bolsters this claim, as it is then argued that η is a 'non-entity' just as in the dynamical approach to SR.

[40] "Recall that in SR, inertial frames are global, which implies that the curvature vanishes everywhere, and hence trivially makes no appearance in the laws of physical interactions" (Brown, 2005, 170f).

[41] This would also support claims to the effect that the spin-2 theory of gravity essentially is nothing but a change of representation (i.e., without new substantial insights, though maybe more apt for unification with other parts of physics; see sections 5 and 6).

[42] In fact, Will (2001) states—though without any further clarification—that all derivations of GR (including the ones we considered) make "implicit use of the SEP" (Will, 2001, 77).

implied by the SEP, then obtaining universal coupling is a trivial result.

However, the general difficulty of relating standard (quantum-)field-theoretic concepts ('coupling') and standard GR concepts ('geometry') is certainly symmetric. Hence, the question whether SEP-GR implies universal coupling seems just as opaque. In fact, the formal requirement of Lorentz invariance in a special-relativistic theory may still be further analysed with respect to its physical implications—for this theory, and (especially) for the full non-linear theory. Again, although the linear theory is manifestly Lorentz-invariant (also locally), and hence in the linear theory the SEP may be considered a presupposition, we should *not* expect this to trivially translate to the SEP holding in the full non-linear theory in general. Depending on how one analyses the SEP, non-linearity may spoil the formerly manifest SEP;[43] after all, the fixed metric η we started from is rendered unobservable (or rather operationally inaccessible) when going to the full theory. But we can indeed demonstrate that a reformulated version of the principle (i.e., universal coupling to g) *must* hold, since in the construction both non-linearity and universal coupling to h are forced on us; and, since g's universal coupling property only depends on the respective property of h due to the reduction, this is preserved in the non-linear theory. Hence, obtaining the reformulated version of the SEP is not a trivial result.

In this sense, only properties of the non-linear theory that are strictly reducible to the respective property manifest in the linear theory (e.g., universal coupling), will hold strictly in the full non-linear theory. Properties of the non-linear theory that are not strictly reducible to the properties manifest in the linear theory (e.g., minimal coupling) will generally not hold strictly, but only effectively (i.e., in the low-energy regime) in the full non-linear theory (something we would expect when going from the linear to the non-linear theory). For low energies we obtain GR (arguably uniquely), while for high energies deviations from GR (modeled as high-energy corrections) are not constrained by the theory.

So, I take it that we *do* learn something substantial from the spin-2 approach: universal coupling may have been a way to think about the SEP all along, but in the spin-2 approach this particular reformulation is derived and shown to strictly translate to the GR case. I propose to read it this way. By establishing universal coupling in the spin-2 derivations from Lorentz invariance, we robustly arrive at a non-geometrical reformulation of the principle in terms of its physical content in the context of (quantum) field theory. This also seems to be in line with physics practice, where non-standard formulations of SEP-GR are rather typical.

Note that the reformulation claim possibly does away with any sharp distinction between WEP-like and SEP-like principles. However, as I shall argue next, it is still possible to distinguish weaker from stronger reformulations of the principle.

So, having established that it is about reformulation, we may ask: why should what has been derived as universal coupling in section 2 be the strongest possible reformulation of some equivalence principle? To begin to answer this question, consider that we may generally posit the following two versions of universal coupling:

UC-Weak The spin-2 coupling to other fields is independent of any of the properties of these fields.

[43] For example, minimal coupling (manifest for the spin-2 case with respect to η) will—in contrast to universal coupling—generally not be manifest for GR with respect to g, as inheriting minimal coupling cannot be secured by the reduction: the other fields are not restricted to couple minimally to h, but only to η.

UC-Strong The spin-2 coupling to all fields (including itself) is independent of the properties of these fields.

Clearly, the second version is stronger than the first.[44] Only UC-Strong implies that gravitational energy is indistinguishable from the energy of ordinary matter fields such that the source for the gravitational field is the total energy-momentum.

To see that it is indeed the strongest version that is derived, we need to appreciate that what has been established by universal coupling in section 2 is not only that the graviton universally couples to ordinary matter, but that it does so to gravitons as well. In particular, self-coupling is *necessary* if any coupling to matter is to be included. Accordingly, WEP-like statements in this context also encompass that "gravitational binding energy of massive bodies will not modify their free-fall motion" (Ortín, 2017).[45] This suggests that the universal coupling reformulation—irrespective of how WEP-like it is—is indeed the most general or strongest possible reinterpretation of the equivalence principles (i.e., UC-Strong). So we may in this sense refer to it as the *strong* equivalence principle. (Again, UC-Strong certainly is one of the requirements for the traditional SEP, but there might be other requirements as well. Accordingly, Ortín (2017) dubs universal coupling a "microscopic version of the SEP".)

After all, the connection between UC-Strong and SEP-GR is also established via the derivation of Einstein's full theory from the spin-2 perspective. In retrospect, from the full derivation we can identify UC-Q and UC-C as similar to SEP-GR in terms of their respective roles. What UC-Strong is for the spin-2, SEP-GR is for non-linear GR.

Let me also briefly comment on why such reformulations are to be expected, especially for quantum-theoretical treatments. Here is what hinders drawing the direct connection from Galileo's original version to the quantum context:

> The Galilean form of the equivalence principle—essentially that all bodies follow the same trajectory when freely falling in a gravitational field with the same initial conditions—is obviously inappropriate when applied to individual leptons. Quantum particles do not follow well defined classical trajectories, and even repeated experiments with a single electron would fail to duplicate the results because of quantum uncertainty.
>
> ...
>
> Faced with this spread in results caused by quantum uncertainty, the best one could do is to define some sort of average trajectory based on expectation values. But even in this quasi-classical sense, a straightforward statement of the principle of equivalence encounters a major obstacle. Fermions, such as electrons, do not follow geodesic paths in a gravitational field, because of the well known spin-curvature coupling ... The spin effect is mass-dependent, which implies that initially coincident electron and muon trajectories (even quasi-classical) will gradually diverge. ... Evidently the most elementary known particles actually fail to comply, even in a

[44] In the literature a variant sometimes appears as the "very strong equivalence principle" (Ciufolini and Wheeler, 1995).

[45] This is another way to make sense of the claim that Schiff's conjecture is supported by the derivations of universal coupling in the spin-2 theories.

statistical sense, with the traditional statement of the equivalence principle. (Davies and Fang, 1982, 470)

Thus, it seems likely that what we obtained (namely UC-Strong) in an explicitly non-geometrical reading of classical field theory and in the manifestly non-geometrical quantum context, can*not* be identified in a strict sense with what is at the heart of the geometrical reading of GR.[46] This, Davies and Fang (1982) continue, then "prompts the search for a more fundamental statement that is directly applicable to fundamental particles. ... A possible statement of the equivalence principle might then be that all fields couple with equal strength to gravity" (Davies and Fang, 1982, 471). This understanding is very much in line with my reformulation claim.

To conclude, employing the symmetry properties of the theory (i.e., global Lorentz invariance of all interactions) has led us to find that the spin-2 field couples universally to all forms of energy-momentum with the same strength (i.e., the coupling of the spin-2 is independent of the properties of the fields it couples to). This can be interpreted as a WEP-like statement. However, since graviton self-coupling is a necessary requirement if any coupling to matter is to be allowed, it is suggested that it is in fact the most general (i.e., the strongest) version of the possible equivalence principles that obtains in this context: a 'microscopic version' (Ortín, 2017) of the SEP.[47] Accordingly, energy-momentum of gravitational fields is indistinguishable from energy-momentum of non-gravitational fields and the source for the gravitational field is the total energy-momentum tensor.

Finally, a few further remarks are in order. In a recent paper, Lehmkuhl (2020) argues that the EEP can be considered a bridge between GR and Newtonian theory, and the SEP can be seen as a bridge from GR to SR. In light of my presentation above, we can now amend this to say that—in some appropriate interpretation of universal coupling and SEP-GR—the SEP indeed remains a bridge between GR and SR, but the direction is turned around: Lorentz invariance—i.e., SR—is where we start, while GR is what we end up with. The SEP is a bridge in the sense that it is where the reduction from GR to a special-relativistic field theory provides explanatory insight. This is a particularly interesting result since the SEP is usually perceived as closely tied to a geometrisation view of gravity, not a 'field only' view (after all, even Brown (2005) employs the SEP precisely to establish chronogeometricity). As I have argued in Salimkhani (2018), the SEP additionally proves to be the 'link' (i.e., bridge) from GR to (quantum-field-theoretic) quantum gravity which underwrites this further.

Lehmkuhl also seeks to understand the role of the principle, and how it constrains the theory (see Lehmkuhl (2020)). From the spin-2 perspective, I take it that we should answer that the focus regarding the conceptual foundations of GR shifts from Lorentz invariance in disguise (i.e., a geometrical version of the SEP) back to the conceptual foundations of SR (i.e., Lorentz invariance proper).

[46] Of course, some (explanatory) connection to geometrical aspects needs to remain: in a dynamical (non-geometrical) understanding of GR the role of SEP-GR is to make sense of chronogeometricity (see also section 4.1).

[47] For example, Nicolis's (2011) assessment is consistent with this. For universal coupling to the (ordinary) matter energy-momentum tensor Nicolis takes "the equivalence principle" (I assume that he refers to the WEP) to be derived. And after including self-coupling so as to obtain coupling to the total energy-momentum tensor, Nicolis explicitly refers to the SEP.

4 Lesson II: A Miracle Disappears

A number of recent works (see Read et al. (2018) and Read (2018; 2019)) have revisited the debate on the standard geometrical approach (GA) versus the dynamical approach (DA) to relativity theory; the dynamical approach was proposed by Brown (2005) and Brown and Pooley (2001; 2006), and generally seeks to argue for what Butterfield (2007) calls "Brown's moral": that chronogeometricity is not a property that geometrical structures have essentially.[48] Since this moral is arguably most convincingly demonstrated in SR in terms of an ontological reduction of the metric to symmetry properties that feature in the matter field dynamics, the doctrine of the dynamical approach may also be considered to be about demonstrating that metrical aspects of spacetime are derivative on matter field dynamics (see Norton (2008)). In the following, I first introduce the respective positions regarding SR and GR. I then address two specific issues that arise within the dynamical approach to GR, which have been dubbed 'two miracles' by Read et al. (2018), and reject the claim that GR faces an unexplained second miracle.

4.1 The Dynamical Approach Revisited

In the case of SR, where we have a fixed background metric (see Pooley (2017)) the two opposing positions are usually characterised as follows:

GA-SR (1) The Minkowski metric field η is ontologically distinct from and independent of matter fields; it is an ontologically primitive entity.
(2) It explains (or determines) the behaviour of matter fields. In particular, it explains why rods and clocks are governed by Lorentz-invariant laws.

DA-SR (1) The Minkowski metric field η is *not* ontologically distinct from and independent of matter fields.
(2) It is a mere codification of the fact that matter field laws are Lorentz (or more precisely: Poincaré) invariant—i.e., η reduces to Lorentz invariance of matter field dynamics.

The geometrical approach to SR (GA-SR) is what can be called the 'orthodox view' (Read et al., 2018) which has metric structure as "a self-standing, autonomous element" (Brown and Pooley, 2006, 84); Friedman (1983), Maudlin (2012), and Norton (2008) hold some version of GA-SR. The dynamical approach to SR (DA-SR), on the other hand, was originally proposed by Brown (2005) and Brown and Pooley (2001; 2006) to give an account of how the metric (or metrical properties) arise in the theory—the issue of chronogeometricity. Regarding this very issue, it is tempting to read GA-SR's second claim as taking chronogeometricity to be an *essential* property of η, i.e., as a necessary property of η in all models of the theory. Note, however, that recent work by Read (2018) significantly refines the typical characterisation of the geometrical approach by distinguishing two variants. I come back to this at the end of this section.

To structure the discussion, it is convenient to distinguish explanatory from ontological issues. Typically, what is called the debate between the geometrical and the dynamical approach is a

[48] Again, chronogeometricity is the property of a field to act as a metric that is surveyed by rods and clocks built from matter fields.

debate about *explanation*. The proposal of the dynamical approach is then often phrased as 'reversing the arrow of explanation' with respect to the geometrical approach, because DA-SR takes the (symmetry properties of the) matter fields to do the explanatory work. Note, however, that DA-SR claims that the Minkowski metric is ontologically reducible to the symmetry properties of matter fields, and, hence, is a 'non-entity' that is erased from the fundamental ontology (see Brown and Pooley (2006)). In this sense, it is misleading to think of DA-SR as simply 'reversing the arrow of explanation'. Especially regarding SR, the dynamical approach expresses a novel *ontological* position. It is opposed to a so-called 'realist'—or rather 'fundamentalist' (see North (2018) and Menon (2019))—conception of spacetime (see Norton (2008)) which, with respect to SR, upholds the distinct ontological status of η in terms of some version of substantivalism.[49] In this sense, DA-SR qualifies as a type of relationalism, as Pooley (2013) argues. To think of DA-SR as an 'anti-realist' conception of spacetime is misleading, as the distinguishing fact between the two rival positions with respect to ontology is not so much a specific stance on realism or anti-realism, but turns rather on the issue of whether spacetime should be considered a *fundamental* structure of the world or not. This issue is addressed towards the end of this paper in more detail.

In the case of GR the debate is different, due to the metric field g being dynamical. In particular, with respect to ontology the proponent of the dynamical approach agrees with the geometrical (or fundamentalist) view that the g field does *not* straightforwardly reduce to matter fields, but is ontologically fundamental, i.e., irreducible to other fields and their properties. It is on the issue of chronogeometricity where the two camps still disagree. Roughly,[50] the geometrical approach takes g to have this property essentially, while the dynamical approach takes g to have this property accidentally. Accordingly, the key issue for the dynamical approach with respect to GR is how the (ontologically distinct) dynamical g field is equipped with its property to act as a metric.

Generally, the two camps maintain the following:

GA-GR (1) The metric field g is ontologically distinct from and independent of matter fields; it is an ontologically primitive entity.
(2) It has its chronogeometric significance essentially, i.e., chronogeometricity is built in as a necessary property of g.

DA-GR (1) The metric field g is ontologically distinct from and independent of matter fields;

[49] A Nortonian spacetime fundamentalist essentially commits to the following claims (cf. Norton (2008)):

(A) There exists a fundamental four-dimensional spacetime for which we can define a set of standard coordinates that are related by certain transformations.

(B) The spatiotemporal interval Δs between two events is a property of spacetime and is ontologically independent of the matter fields.

(C) Rods and clocks survey spacetime (i.e., measure spatiotemporal distances) because they are built by matter fields that obey laws that are adapted to the spacetime geometry (which is ontologically independent of the matter fields).

Just as mentioned above for GA-SR's second claim, some further explication is also needed for Norton's notion of 'adaptation' which is essential for the concept of chronogeometricity. In this paper, I mostly use Read et al.'s (2018) notion of symmetry 'coincidence'.

[50] Again, a qualification does apply due to Read (2018) and will be discussed in a moment.

it is an ontologically primitive entity. Contrary to DA-SR, g is *not* reducible to symmetry properties of matter fields.
(2) Chronogeometricity is *not* an essential property of g, but accidental. Chronogeometricity is (contingently) acquired by means of the (empirical fact of the) SEP.

So in contrast to DA-SR—where chronogeometricity follows automatically from the metric field being a mere codification of the symmetry properties of matter field dynamics—DA-GR needs to invoke an additional postulate that is put in by hand, namely the SEP.[51] For the dynamical approach, it is via the SEP that g acquires its chronogeometric significance.[52] This is because "nothing in the form of the equations *per se* indicates that $g_{\mu\nu}$ is the metric of space-time" (Brown, 2005, 160), rather "it is because of ... local Lorentz covariance that rods and clocks, built out of the matter fields which display that symmetry, behave as if they were reading aspects of the metric field and in so doing confer on this field a geometric meaning" (Brown, 2005, 176).[53]

On the other side, Read et al. (2018) take proponents of GA-GR typically to argue that "the metric field has a primitive connection to spacetime geometry". It is "the existence of the Lorentzian metric field" which "*explains* the form of the local dynamical laws in the theory" (Read et al., 2018, 19). The explanatory work in GA-GR is done by the fundamental g field alone. Since g is argued to have its chronogeometricity necessarily in GA-GR, the SEP is not considered to do any additional explanatory work.[54]

Thus, ontologically, the situation is the same in both accounts, but there is a difference regarding the explanatory aspects. In GA-GR, the g field has its chronogeometricity essentially. Therefore, g is viewed as straightforwardly explaining the form of the dynamics. In DA-GR the explanation does *not* run straightforwardly, but via the postulated SEP (Brown, 2005, 151)—since the g field does *not* have chronogeometricity essentially, but needs to acquire it first. This conceptual approach is similar to that preferred by particle physicists like Steven Weinberg (see Weinberg (1972)): that is to say, starting from some physical ground—i.e., the SEP as the (empirical) core principle that captures what gravity is about—rather than taking geometry straightforwardly to feature in the explanation.

Adopting this approach arguably brings with it the task of explaining—or, at least, making plausible—why the SEP should hold in the first place (see also Salimkhani (2018) and Weinberg (1972)). Otherwise the SEP is *just* a brute empirical fact; and explaining chronogeometricity by a brute fact might be considered just as dissatisfactory as GA-GR's essentialism (where the g field *just has* this property). So, without making it at least plausible why the SEP should hold, DA-GR runs the risk of being unable to fully undermine the geometrical view (which, in a sense, is free of this explanatory burden).[55]

[51]The SEP does not follow from the Einstein field equations, but needs to be introduced by hand (e.g., Brown (2005) and Lehmkuhl (2020)).

[52]Read (2018) raises doubts that Read et al. (2018) are correct in taking the SEP as being a necessary condition for chronogeometricity. Indeed, there may be other ways to gain "operational access to the metric field" (Read, 2018)—notably explored by Ehlers et al. (1972).

[53]See also Read (2018).

[54]Quite the opposite, as it can be argued that the SEP is explained by this (see also below).

[55]Essentialising chronogeometricity in GA-GR arguably provides an explanation for why the empirical fact of the SEP holds: if the Lorentzian metric field has its chronogeometric property essentially, then the SEP comes for free (i.e., holds necessarily), and "the existence of the Lorentzian metric field explains the form of the local

Indeed, DA-GR as it stands does not offer a more substantial explanation for why the SEP and, hence, chronogeometricity holds; this is expressed more precisely in the two 'miracles' which I shall come to in a moment. It is by raising certain problem cases for the geometrical approach that Read et al. (2018) and Read (2018) arguably successfully balance this explanatory deficit of the dynamical approach. In particular, Read (2018) refines this rationale and significantly clarifies the debate by, first, distinguishing two versions of the geometrical approach—the qualified geometrical approach (QGA), and the unqualified geometrical approach (UGA);[56] second, arguing that only QGA is tenable due to such problem cases; and, third, arguing that a tenable (i.e., qualified) version of the geometrical approach and the dynamical approach are explanatorily on a par (i.e., both must accept the same brute facts). In fact, Read (2018) concludes that any distinction between the dynamical and the tenable geometrical approach to GR collapses.

For SR the situation is different: the dynamical approach and the qualified geometrical approach are distinguished, on the one hand, by their ontological claims, and, on the other hand, by the fact that the dynamical approach—unlike the qualified geometrical approach—rigorously explains chronogeometricity (with accepting less brute facts) via the ontological reduction of the metric field (see also Read (2018)). I come back to this in a moment when discussing Read et al.'s (2018) two 'miracles'.

But first let me draw attention to the fact that for both the classical and the quantum spin-2 theory (modulo the additional field) the situation is that of SR. In particular, the geometrical and the dynamical approach differ with respect to both ontology and explanation. Ontologically, the geometrical approach to spin-2 theory postulates the Minkowski metric and the matter fields (including the additional graviton field), whereas the dynamical approach postulates only the matter fields (including the additional graviton field). The explanatory work in the geometrical approach is done by the Minkowski metric, whereas in the dynamical approach the explanatory work is done by the symmetry properties of the matter field dynamics (including the graviton field). As will become apparent below, the dynamical approach to spin-2 theory is preferable to the geometrical approach with respect to ontological commitments and explanatory robustness and rigour. Hence, it is the spin-2 approach that fully delivers on the explanatory promise of the dynamical approach and distinguishes it from other positions.

dynamical laws in the theory" (Read et al., 2018, 19).

[56] Both QGA and UGA are geometrical approaches in the sense that they affirm the ontologically distinct status of the metric field in the respective theory. With respect to explanation, UGA claims that the metric field of some theory (e.g., SR) self-standingly explains the dynamical symmetries of all possible dynamical laws of that theory without any further assumptions. On UGA, the metric is understood to have chronogeometricity essentially (or necessarily), i.e., in all models of the theory without any restriction. On the other hand, QGA claims that—given a particular dynamical equation—the dynamical behaviour of a certain set of matter fields may be explained by the metric field because in the particular dynamical equation under consideration that set of matter fields couples to the metric field. On QGA, the metric has chronogeometricity only under certain restricting conditions for the particular dynamical equation. Therefore, QGA can be considered to accept the chronogeometricity charge by the dynamical approach (Read, 2019, 2), i.e., QGA accepts that chronogeometricity is, in a sense, an accidental property of the metric field. As Read (2018) remarks, it is primarily UGA that is attacked by Brown (2005), Brown and Pooley (2001; 2006), and Read et al. (2018).

4.2 Just One Miracle in GR

The issue that trying to explain chronogeometricity potentially involves accepting certain unexplained brute facts can be made more precise by distinguishing between: (1) the fixing of a universally shared dynamical symmetry property for all matter fields (expressing the core statement of the SEP), and (2) the coinciding of this shared dynamical symmetry property with the symmetry property of the metric field (thereby conveying chronogeometric significance to the metric field). Accordingly, Read et al. (2018) identify the following two potential 'miracles' (i.e., unexplained brute facts) in the foundations of relativity theory:[57]

MR1 All matter fields exhibit local Poincaré invariance.

MR2 The local symmetry properties of matter fields coincide with the local symmetry properties of the dynamical metric field.

For SR, the dynamical approach accepts only MR1 as unexplained—indeed, MR2 receives an explanation through the ontological reduction of the metric field. For GR, the dynamical approach accepts both MR1 and MR2 as unexplained brute facts. In contrast, the geometrical approach (or, more precisely, its unqualified variant[58]) "attempts to rationalise" (Read et al., 2018, 20) both miracles by means of the respective ontologically primitive metric field which is argued to appropriately 'constrain' the dynamics of the matter fields, i.e., to explain the miracles (see also Read (2018)). However, such an explanation must seem question-begging from the perspective of the dynamical approach, or so Read et al. (2018) argue:

> One thing that the advocate of (A) [the (unqualified) geometrical approach] may say here is the following: Minimally coupled dynamical laws in GR feature the metric field g_{ab}; as we have seen, the presence of (or rather, the coupling to) *this metric* constrains the local form in the neighbourhood of any $p \in M$ of the dynamical laws of those fields to which it couples. Consequently, the symmetries of the local dynamical laws *must* coincide with the symmetries of the metric field. This argument misses the point, however, for the very issue in question is why the dynamical laws governing matter fields take such a form—rather than another, with *different* local symmetry properties. In other words: *why this particular coupling?* This is the essence of MR2, which remains untouched by such arguments. (Read et al., 2018, 20)

The concrete problem cases that Read et al. (2018) present demonstrate that the explanatory work cannot be done by the metric field alone—contrary to what is claimed by the (unqualified) geometrical approach. Absent further argument, the position is indeed not tenable.

Read (2018) agrees that the attacks on what he dubs the unqualified geometrical approach do find their mark: UGA is to be dismissed on the basis of the problem cases (see Read et al. (2018) and Read (2018)); without further input assumptions, the metric field g does not suffice to obtain MR1 and MR2. It just is perfectly consistent with SR or GR that the matter field

[57]Instead of "non-gravitational fields", I shall use "matter fields" in this paper. This is to anticipate the newly introduced gravitational 'matter' field h in the spin-2 approach.

[58]Again, it is the unqualified variant of the geometrical approach with which the works prior to Read (2018) primarily take issue.

dynamics do not manifest the symmetries of the respective metric field. For instance, without further input assumptions "there is nothing to rule out dynamical equations for non-gravitational fields in GR which are locally *Galilean* invariant" (Read, 2019, 3).

This is accepted by any tenable variant of the geometrical approach (i.e., QGA). It is only with the help of further input assumptions, which restrict the possible dynamics of matter fields, that the metric field can feature in explanations of matter field dynamics. Accordingly, QGA accepts both miracles as unexplained brute facts—not only in GR, but in SR as well (Read (2018); see also below).

According to Read et al. (2018) and Read (2018), this all suggests the following conclusion: the unexplained existence of MR1 and MR2 is *independent* of endorsing a particular view in the debate. It is a fact about GR itself (not about a particular interpretation) that MR1 and MR2 do not receive an explanation. Regarding MR2, this is due to the apparent irreducibility of g:

> if one could argue that what had previously been regarded as the ontologically independent metric field of GR was in fact reducible to a codification of symmetries of the dynamical equations of matter fields (as ... in SR), then this would provide an explanation for MR2. (Read et al., 2018, 20)

This is why Read (2019) turns to string theory.

But is this really a deficiency of GR that may only be resolved in a successor theory? The results in section 2 put some pressure on this conclusion. Classical spin-2 theory, which is equivalent to GR (up to a few qualifications), already provides what is required: a reconceptualisation of GR that—in a dynamical reading—allows for an ontological reduction of GR's metrical structure to matter field structure. Hence, I take it that MR2 is *not* generally miraculous in GR, but only within a (tenable) geometrical interpretation.

In a first step, recall that for the spin-2 theory itself the situation is just that of SR (modulo the additional matter field). Accordingly, the dynamical approach to spin-2 theory takes the symmetry properties of η to necessarily (or, automatically) coincide with the dynamical symmetry properties of the matter fields (including h) due to an ontological reduction of η to matter fields *à la* the dynamical approach to SR (η is nothing but a codification of the universally shared dynamical symmetry properties). It is only the specific type of dynamical symmetry properties of the matter fields that is contingent (and is to be determined empirically: it happens to be an empirical fact that the matter field laws are Lorentz-invariant). The dynamical approach to spin-2 theory takes MR1 as a brute fact and explains MR2.

In a second step, one then essentially appeals to the fact that classical spin-2 theory and GR are equivalent:[59] independently of endorsing a particular position in the debate on the dynamical approach, g (in the full non-linear theory) is no longer taken to be ontologically distinct and autonomous (as both positions standardly have it for GR), but is taken to be ontologically

[59] If one takes the qualifications that apply to this equivalence claim seriously, classical spin-2 theory is not equivalent to standard GR in the strictest sense possible, and, hence, the lesson drawn in this paper regarding MR2 has to be slightly weakened in general. This is because the validity of the ontological reduction of g to the ingredients of the spin-2 theory rests on the two theories being equivalent. In fact, if one could show that there are qualifications to the equivalence claim which specifically restrict the theory by effectively excluding potential problem cases that violate MR2—similar to the additional input assumptions in QGA—one may argue that, after all, the spin-2 view does *not* explain MR2. Then MR2 really would remain unexplained in GR.

reducible to η and h.[60] Consequently, from the perspective of a dynamical approach, g inherits the shared dynamical symmetry properties of all the matter fields (including h), and the universal coupling property of h. Accordingly, on a dynamical reading, the symmetry properties of g necessarily coincide with the dynamical symmetry properties of the matter fields (including h), and g acquires its chronogeometric significance. It is only the specific type of the dynamical symmetry properties of matter fields that is contingent. The dynamical approach to GR based on the spin-2 theory has to take MR1 as a brute fact, but is able to explain MR2. In light of this argument, the second miracle only seems to be an unexplained brute fact in GR due to an obscuring representation of the theory—which, ironically, is the standard representation, of course.

At the heart of this 'spin-2 argument' lies the fact that it is actually possible (*pace* the previous verdicts on the dynamical approach to GR) to separate the dynamical and fixed parts of g in order to obtain a fixed-field formulation of GR. Generally, fixed-background theories allow for an ontological reduction in which MR2 does not appear miraculously, whereas fully dynamical theories do not (see Read (2018)). In this sense, the success of the spin-2 argument can be anticipated (which does not render it uninteresting, of course). The following is important to note, though: if the two formulations are indeed equivalent (and do not actually represent different theories), all conceptual aspects should be fully translatable. In particular, there should be a fact about whether GR is fully dynamical or has a background. The solution to this is to appreciate that the fixed part is redundant to specific symmetry properties; fixed fields are not ontologically independent. I shall briefly come back to this at the end of next section.

I take it that, in light of the presented argumentation, the dynamical approach to (a spin-2 understanding of) GR is preferred over the qualified geometrical approach to (a spin-2 understanding of) GR, for essentially the same reasons that Read (2018) puts forward in the case of SR: the dynamical approach is preferred because (a) the dynamical approach only has to accept MR1 as unexplained, while QGA has to accept both miracles;[61] and (b) the dynamical approach has a 'simpler' ontology than QGA (see also section 5). GR understood in light of the spin-2 approach demonstrates that local symmetry properties of matter fields do not miraculously coincide with the symmetry properties of the g field. In a sense, MR2 is reduced to MR1.

5 Lesson III: The Dynamical Approach Resurrected

Let us briefly retrace our steps. We have seen that it is possible to arrive at full non-linear GR with its dynamical (curved) metric field g from a (classical or quantum) field theory of a massless spin-2 field on a fixed (flat) Minkowski background. At heart, GR still is a special-relativistic theory. The spin-2 view suggests that the hitherto ontologically fundamental metric field g is, in fact, non-fundamental. We can understand this as follows: g is ontologically reduced to two fundamental entities, the fixed Minkowski field η and the dynamical spin-2 field h. This amounts

[60] This is based on (a) g being obtained as the effective metric field by reexpressing the sum of η and h in the derivation of the EFE (see section 2.1)—or, less formalistically, η and h adopting the functional roles of g—and (b) the equivalence claim.

[61] QGA takes the metric η as ontologically autonomous. Therefore, the only possible explanation of MR2—reducing the metric to symmetry properties of matter field dynamics—is not available.

to a geometrical interpretation, analogous to the (qualified) geometrical approach to SR (see Read (2018)).

However, an ontologically more radical and conceptually more coherent[62] reading of the non-fundamentality of g is almost self-evident. Recall that fixed background theories (like SR) allow for an ontological reduction of the fixed background to matter field dynamics: the dynamical approach to SR ontologically reduces the Minkowski metric η to the symmetry properties of the dynamics of matter fields. Accordingly, the Minkowski metric η does not appear in the fundamental ontology of SR, but is considered a "glorious non-entity" (Brown and Pooley, 2006): η is a mere codification of the universally shared dynamical properties of the matter fields.

This can be used to reinterpret the non-fundamentality of g, and, thereby, provide a fully dynamical picture of GR: the fixed background that features in the spin-2 theory (the Minkowski metric η) is—just as in the case of SR—merely a codification of the symmetry properties of the matter field dynamics (now also including h). Thus, we can go beyond reducing g to η and h (as the geometrical view has it). A dynamical understanding of the spin-2 view of GR suggests that the hitherto primitive metric field g is ontologically reduced to Lorentz-invariant matter field dynamics. Both the chronogeometric and the gravitational significance[63] of g are accounted for by this ontological reduction: g inherits the properties of h and obtains its capacity to act as the spacetime metric. We are left with a position that only commits to the fundamentality of matter fields, i.e., arguably some version of relationalism (see Pooley (2013)). The dynamical approach to GR is resurrected[64] by rendering (the metrical aspect of) spacetime non-fundamental (or, non-substantival).

Note that there remains a sense in which a substantivalist view is feasible: the universally coupled h field certainly is substantival and may—due to its properties—be considered spacetime (cf. Knox (2019)). However, this is arguably different from what we have according to typical versions of spacetime substantivalism (especially with respect to the quantum spin-2 theory). Most importantly, the operationally accessible spacetime—or rather, its metrical aspect—effectively arises from interacting (quantum) matter fields. This contributes significantly to giving up any sharp distinction between matter and spacetime, thereby supporting a recent proposal by Martens and Lehmkuhl (2020*a*; 2020*b*).

But there is another sense in which substantivalism is not ruled out. The dynamically interpreted spin-2 theory suggests an understanding of GR spacetime as a *partly* non-fundamental structure: as presented, it does away with the *metrical* properties of spacetime[65] by reducing

[62] For instance, regarding the number of unexplained brute facts that are accepted.

[63] According to Lehmkuhl (2008), "a mathematical object has gravitational significance if it plays a role in describing and explaining the phenomena we count as gravitational".

[64] Again, the significance of this feature of the spin-2 approach—and, specifically, endorsing the image of the dynamical approach being resurrected—depends on what one takes the central claim of the dynamical approach to be. If one thinks that the dynamical approach is mainly a position on chronogeometricity (Brown and Read seem to hold this view), the dynamical approach to GR can be considered to not be stuck in the first place and, accordingly, the resurrection image might seem misleading. This is a viable position, but—*pace* this—I take it that DA-SR's ontological reduction claim is more important. This is because the reduction claim is (a) stronger, and (b) provides a more robust and more rigorous explanation of chronogeometricity (since chronogeometricity then follows automatically without MR2 appearing miraculously).

[65] Recall that spacetime is represented by the tuple of two mathematical objects, a metric and a manifold. We can understand this as ontologically committing us to two fundamental entities in the world (represented by these two mathematical objects). Or we may equally well say that the entity *spacetime* has (at least) two aspects, or

them to certain properties of matter fields; thus, metrical properties are not primitive, not fundamental, and not independent. What remains untouched, however, are the *topological* properties of spacetime, i.e., the manifold. I will briefly comment on this issue in section 6.

Let me stress that I generally take it that the foundational and philosophical importance of the spin-2 approach is not tied to the dynamical interpretation. Still, it is the dynamical perspective that provides the ontologically most parsimonious and conceptually most coherent take on the spin-2 approach to GR. In particular, the dynamical approach perspective helps to resolve potential reservations against the spin-2 view on GR—and vice versa.

For instance, Pitts (2016) raises doubts about claims to the effect that the ontology of the spin-2 approach is obvious, and that the derivation shows that GR simply becomes another special-relativistic field theory. Pitts argues that it is not clear whether the unobservability (e.g., Thirring (1961)) of the flat metric, from which one starts in the spin-2 approach, accounts for it 'not being real' (which Pitts takes to be tacitly assumed), and, hence, for it 'ceasing to exist' (Pitts, 2016). However, in the dynamical perspective, the flat background is a non-entity from the start—it does not 'cease to exist'. More importantly, its being a non-entity is not justified by its unobservability, but by its being *redundant* (it is just a codification of the dynamical symmetry properties of the matter fields). It is key to note that this argument is not available for a geometrical reading of the spin-2 approach to GR.

All this is not sufficiently appreciated by Brown (2005), who only briefly mentions the possibility of a spin-2 derivation, without properly acknowledging the strength of the result or its usefulness for his programme. Brown notes that "[i]t is widely known that there is a route to GR based on global Minkowski spacetime with spin-2 gravitons" (Brown, 2005, 172), but (1) takes it that this involves 'requiring' "that $h_{\mu\nu}$ couple to its own stress-energy tensor, as well as to the matter stress-energy tensor" (Brown, 2005, 172). This downplays the most significant aspect of the derivation, namely that the coupling properties of h result from merely demanding mathematical consistency. For consistency reasons, an interacting spin-2 theory is forced to reproduce GR. Furthermore, Brown (2) does not make it explicit that this may be read as providing a fixed-field formulation of GR that then allows for similar ontological conclusions as in the case of SR; and, (3) does not discuss the issues that arise when taking this 'route to GR' seriously (e.g., qualifications regarding the equivalence of the theories). Brown does point out, though, that "[t]he flat background metric in this approach becomes bereft of any direct operational meaning because the SEP holds for the local inertial frames defined relative to $g_{\mu\nu}$ and not to $\eta_{\mu\nu}$" (Brown, 2005, 172). However, in line with my second comment, this does not give sufficient credit to the core idea of the dynamical approach to special-relativistic scenarios: the flat background field's being a non-entity.

Before I turn to a few critical remarks regarding, for example, the ontological inferences, note that these results may additionally be viewed to have impact on the issue of *background dependence* usually raised for such fixed-background theories. In brief, it is argued that one of the important features of GR is that the metric field g is dynamical, hence the theory is background-independent.[66] Both spin-2 theories, on the other hand, apparently have a fixed background, namely the Minkowski metric, which may be taken to render the theory inferior

properties, namely the metrical and topological ones.

[66]For a proper discussion of this issue, the reader is referred to Belot (2011) and especially Read (2016).

to GR.[67] However, the proponent of a dynamical approach to GR may argue that this 'fixed' background is a mere codification of an empirical fact: Lorentz symmetry. Its being 'fixed' is just the first miracle: it is a brute empirical fact that all laws of physics have this symmetry property.[68]

6 Conclusion and Critical Remarks

In this paper, I have argued that three lessons can be drawn from the spin-2 approach: (1) the full non-linear theory of GR is derivable in classical and quantum spin-2 theory, and in particular the equivalence principle can be shown to reappear as 'universal coupling'; (2) appreciating the insights from classical spin-2 theory, the so-called second miracle is not miraculous, but explained; (3) the results from the spin-2 theories allow for a full resurrection of the dynamical understanding of spacetime in GR. In particular, we see that key aspects of spacetime (metrical aspects) dissolve in GR, thus foreshadowing current developments in the context of (speculative) theories of quantum gravity.

Finally, a few potential criticisms of my approach need to be addressed. First, I have argued that the Lorentz invariance of the fields is unexplained and needs to be accepted as a brute fact (MR1). Furthermore, the brute empirical fact of Lorentz invariance turns out to be the essential ingredient for the argument I present. Now, one might ask whether taking Lorentz invariance as the empirical basis for the foundational issues tackled here is actually sufficiently justified. What, one might ask, if Lorentz invariance turns out to hold only effectively, i.e., is violated in the fundamental theory? In fact, many theories of QG require Lorentz invariance to be violated; however, regarding empirical data, no violations of Lorentz invariance have yet been found (see Mattingly (2005)). So any violation of Lorentz invariance is expected to be a high-energy effect. As Carlip (2014, 203) points out, "even violations at very high energies can feed back into quantum field theory through loop effects and lead to drastic consequences at low energies". Thus, the empirical results for Lorentz invariance are further strengthened. What is more, small violations of Lorentz invariance are considered to "lead to problems with black hole thermodynamics" (Carlip, 2014, 203).

Second, with only the metrical aspects of spacetime being reduced to symmetry properties of matter fields, one problem for a full-blown constructive understanding of GR (and spacetime) remains unsolved: manifold structure still seems to be presumed.[69] This leaves us with the so-called problem of pregeometry that Norton (2008) is also concerned with (for recent work on this topic see Menon (2019), Stevens (2018), and Linnemann and Salimkhani (2020)). Prominently, Norton observed that even the dynamical approach to SR (which is of the same type as the dynamical approach to spin-2 theory) is at best 'half-way'. Thus, the problem of pregeometry is also acute for the spin-2 approaches. The asserted reduction of GR metric structure to

[67] There are several caveats to this. First and foremost, one needs to properly define the notion of background independence; depending on the notion used, GR may or may not be background-independent (see Read (2016)). Also, this depends on whether spacetime is considered to be empty (e.g., Belot (2011)).

[68] There is another aspect to the issue of background independence: one could choose another (non-Minkowskian) background. In other words, one could choose a different 'split' of g—i.e., $g = \tilde{g} + \tilde{h}$ (see section (2.1)).

[69] Spacetime is usually identified with the pair $\langle M, g \rangle$ of manifold structure M and metric structure g.

certain symmetry properties of matter fields might be considered successful, but the manifold structure is tacitly presumed. In a sense, we here observe a push-back to refocussing on manifold structure, something which got pushed aside in the famous debate on the hole argument. It is here where one also sees most clearly that the issue of the dynamical approach is more closely tied to the traditional debate on substantivalism than is usually acknowledged. However, regarding the problem of pregeometry this can still be considered progress, because 'half-way constructivism' is also an interesting explanatory alternative (see Pooley (2013)). Its ontology is less structured (than brute substantivalism), and as argued by Menon (2019) and Linnemann and Salimkhani (2020), this may be advantageous with respect to future (quantum-theoretical) spacetime theories.

Third, why should unification attempts like Weinberg's be considered to have ontological import if the resulting theory is only an effective theory? With Weinberg (1999), we may reply as follows. In the modern effective field theory view, physical theories are typically not straightforwardly fundamental. Hence, we should be honest about what physics can and does provide us with:

> I think that in regarding the standard model and general relativity as effective field theories we're simply balancing our checkbook and realizing that we perhaps didn't know as much as we thought we did, but this is the way the world is and now we're going to go on the next step and try to find an ultraviolet fixed point, or (much more likely) find entirely new physics. (Weinberg, 1999, 250)

But—*pace* Redhead (1999)—this does not mean that effective field theories thereby block the search for fundamental laws of nature or scientific realism (see Williams (2017)). It may seem no surprise, though, that probably the most important reductive aspect of the spin-2 proposal concerns the reduction of a theoretical concept—the SEP—to another (arguably more fundamental) theoretical concept—Lorentz invariance. It is this conceptual reduction that—in light of the dynamical approach—turns out to fuel a reduction in terms of ontology (cf. Morrison (2006) claiming that physics is usually about reduction of theoretical principles, not ontological entities).

Fourth, why should one commit to this 'ugly way' of thinking about GR, especially when, in the end, it gives rise to g? First of all, the spin-2 approach needs to give rise to g (at least approximately), otherwise it would (empirically) fail. Secondly, rather than getting distracted by notoriously vague concepts like 'beauty', we should appreciate that it is about issues of, on the one hand, (direction of) explanation, and, on the other hand, fundamentality (or ontological dependency). When it turns out that g may be viewed to 'emerge' (to use another notoriously vague notion in a loose sense for a moment) from some *more generic* structures, namely matter fields, this should be taken seriously. (There is no doubt that the objects of statistical mechanics are more fundamental than the objects in thermodynamics, although thermodynamics may often provide the more compact or—if one wishes—'beautiful' description.)

Hence, one reason to prefer the spin-2 view over the standard view is that it minimises ontological commitments (not merely quantitatively, but qualitatively: the commitments are less specific, more generic).[70] However, and here is the fifth objection, why should we take this

[70] Also, the spin-2 view may be (empirically) preferred due to its feature of being arguably more restrictive

as suggesting an ontological reduction in the first place? Why is this not merely a change of representation? Or similarly, how should we decide which of the pictures is true of the world?[71]

Indeed, it is crucial that the classical spin-2 approach on its own is not conclusive with respect to reevaluating ontology. To make an ontologically clear case for g being non-fundamental, we would need to argue that the properties of g are determined by the properties of η and h—*but not vice versa*. One may raise severe doubts about whether this is possible. This is because of the following symmetry: due to $g = \eta + h$, g may be viewed as determined by η and h, but in the same sense, h may be viewed as determined by g and η. Just splitting g into 'parts' does not do the job. In a sense, 'parts' and 'whole' determine each other mutually. We may have established some ontological dependence between g, η, and h, but we have not yet established any direction for this dependence. In particular, we may equally well argue that it is g which is ontologically fundamental; mathematically, we can just run the whole derivation backwards. This is most notably the case for the geometrical view: it seems that when considering both h and η as self-standing 'parts' of g, there is no merit at all in splitting g with respect to explanatory and ontological considerations (see also the discussion of the two miracles in section 4.2).[72] This arguably changes when adopting the dynamical approach. Here, a more coherent perspective is available that—similar to my response to the fourth objection—might explain why h should be considered fundamental, and why g should be considered derivative: namely, because η is a non-entity, redundant to certain symmetry properties of the matter field dynamics (including the dynamics of h). Hence, there is a sense in which the ontological dependence of h and g is not analogous to a simple part-whole dependence. Still, why is this not merely a change of representation?

We here face a typical underdetermination problem, or selection problem, in metaphysics. How might this be solved? As I have recently argued in Salimkhani (2019), unification might be a way out. If we understand unification as a *result* of science rather than an imposition on (or aim of) science, we may hope that a physical theory of sufficiently large scope will prefer one ontology over the other. A limited scope theory may allow for several conflicting ontologies, but widening the scope then results in constraints. An immediate idea would be that the alternative view on GR is better suited for a unification with particle physics (the other pillar of modern physics besides GR). So, particle physics would be used as a continuity condition for our interpretation of GR. In fact, we may even argue that we already have a working quantum theory of gravity in terms of the quantum spin-2 approach (see Salimkhani (2018)). The problem, infamously, is that

with respect to topology or causality. However, it is important to note that keeping in mind some central features of GR (e.g., the singularity theorems), one might push against such claims.

[71] In fact, Sexl and Urbantke (2002) conclude that these findings confirm the famous charge of conventionalism raised by Poincaré against GR. Note, however, that Poincaré did not argue that we cannot settle for what is true of the world, but that such findings indicate that there is no such thing as spacetime geometry. Our being free to choose a representation conventionally is taken to prove that it really is nothing but a convention with no ontological dimension whatsoever. While this latter historical position adopted by Poincaré might be avoidable, the former version is not (at least with respect to classical GR on its own).

[72] Actually, the issue is even worse: we could also use a different background, i.e., we could 'split' g differently. The choice of one particular background, namely η, might then even be viewed as bringing in MR2—notably, without explanation (*pace* my previous claim). However, we can still appeal to the following explanation: distinguishing the choice of η from other choices is not an arbitrary posit, but is explained by its providing a best systematisation—especially with respect to the dynamical approach.

these arguments are in a sense undermined by the fact that QFT gravity is non-renormalisable, and therefore considered untenable as a fundamental theory (though still fine as an effective theory). However, that we have not yet solved these issues does not temper the general rationale of constraining ontology via unification. And, at least at the moment, the spin-2 view seems to be preferred, because it also fares well with our most promising candidate for a theory of QG, namely string theory.

Acknowledgements. I would like to especially thank Andreas Bartels, Harvey Brown, Stanley Deser, Juliusz Doboszewski, Jamee Elder, Nick Huggett, Dennis Lehmkuhl, Niels Linnemann, Niels Martens, Tushar Menon, Brian Pitts, James Read, Christian Röken, Thorsten Schimannek, and two anonymous referees for very helpful discussions and highly relevant feedback on earlier versions of this paper. Furthermore, I thank the audiences of the *British Society for the Philosophy of Science Annual Conference (BSPS 2017)* in Edinburgh, the symposium 'Promoting the Field View of General Relativity: New Insights From Spin-2, Emergent Gravity, and the Dynamical Approach' at the *6th Biennial Conference of the European Philosophy of Science Association (EPSA17)* in Exeter, the international workshop *Spacetime: Fundamental or Emergent?* in Bonn in October 2017, and the *History and Philosophy of Physics Research Seminar* in December 2019 in Bonn for useful comments.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), research group *Inductive Metaphysics*, grant number FOR 2495; and by the Doctoral Programme of the Faculty of Arts at the University of Bonn.

References

Alvarez, E. (1989), 'Quantum gravity: an introduction to some recent results', *Reviews of Modern Physics* **61**(3), 561–604. DOI: 10.1103/RevModPhys.61.561.

Arnowitt, R., Deser, S. and Misner, C. W. (2004), 'The dynamics of general relativity', arXiv: 0405109 [gr-qc]. Originally appeared in L. Witten (Ed.), *Gravitation: An Introduction to Current Research*, Chapter 7, pp. 227–264. New York: Wiley, 1962.

Balashov, Y. and Janssen, M. (2003), 'Presentism and relativity', *The British Journal for the Philosophy of Science* **54**(2), 327–346. DOI: 10.1093/bjps/54.2.327.

Belot, G. (2011), 'Background-independence', *General Relativity and Gravitation* **43**(10), 2865–2884. DOI: 10.1007/s10714-011-1210-x, arXiv: 1106.0920 [gr-qc].

Boulware, D. G. and Deser, S. (1975), 'Classical general relativity derived from quantum gravity', *Annals of Physics* **89**, 193–240. DOI: 10.1016/0003-4916(75)90302-4.

Brown, H. R. (2005), *Physical relativity: Space-time structure from a dynamical perspective*, Oxford University Press, Oxford.

Brown, H. R. and Pooley, O. (2001), The origins of the spacetime metric: Bell's Lorentzian pedagogy and its significance in general relativity, *in* C. Callender and N. Huggett, eds, 'Physics Meets Philosophy at the Plank Scale', Cambridge University Press, Cambridge, pp. 256–272.

Brown, H. R. and Pooley, O. (2006), Minkowski space-time: A glorious non-entity, *in* D. Dieks, ed., 'The Ontology of Spacetime', Vol. 1 of *Philosophy and Foundations of Physics*, Elsevier, Amsterdam, chapter 4, pp. 67–89.

Brown, H. R. and Read, J. (2018), 'The dynamical approach to spacetime theories', PhilSci-Archive: 14592. Forthcoming in E. Knox and A. Wilson (Eds.), *The Routledge Companion to Philosophy of Physics*. London/New York: Routledge.

Butterfield, J. (2007), 'Reconsidering relativistic causality', *International Studies in the Philosophy of Science* **21**(3), 295–328. DOI: 10.1080/02698590701589585.

Carlip, S. (2014), 'Challenges for emergent gravity', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **46**, 200–208. DOI: 10.1016/j.shpsb.2012.11.002.

Carroll, S. (2004), *Spacetime and Geometry: An Introduction to General Relativity*, Addison Wesley, San Francisco.

Ciufolini, I. and Wheeler, J. A. (1995), *Gravitation and Inertia*, Princeton University Press, Princeton.

Coleman, S. and Mandula, J. (1967), 'All possible symmetries of the S matrix', *Physical Review* **159**(5), 1251–1256. DOI: 10.1103/PhysRev.159.1251.

Davies, P. C. W. and Fang, J. (1982), 'Quantum theory and the equivalence principle', *Proceedings of the Royal Society of London A* **381**, 469–478. DOI: 10.1098/rspa.1982.0084.

Deser, S. (1970), 'Self-interaction and gauge invariance', *General Relativity and Gravitation* **1**, 9–18. DOI: 10.1007/BF00759198.

Deser, S. (1987), 'Gravity from self-interaction in a curved background', *Classical and Quantum Gravity* **4**(4), L99–L105. DOI: 10.1088/0264-9381/4/4/006.

Deser, S. (2010), 'Gravity from self-interaction redux', *General Relativity and Gravitation* **42**, 641–646. DOI: 10.1007/s10714-009-0912-9.

Deser, S. and Henneaux, M. (2007), 'A note on spin-2 fields in curved backgrounds', *Classical and Quantum Gravity* **24**, 1683–1685. DOI: 10.1088/0264-9381/24/6/N01.

Di Casola, E., Liberati, S. and Sonego, S. (2015), 'Nonequivalence of equivalence principles', *American Journal of Physics* **83**(1), 39–46. DOI: 10.1119/1.4895342. arXiv: 1310.7426 [gr-qc].

Donoghue, J. (1994), 'General relativity as an effective field theory. The leading quantum corrections', *Physical Review D* **59**, 3874–3888. DOI: 10.1103/PhysRevD.50.3874.

Ehlers, J., Pirani, F. A. E. and Schild, A. (1972), The geometry of free fall and light propagation, *in* L. O'Reifeartaigh, ed., 'General Relativity: Papers in Honour of J. L. Synge', Clarendon Press, Oxford, pp. 63–84.

Einstein, A. (1914), 'Principles of theoretical physics', Inaugural address before the Prussian Academy of Sciences. First published in the Proceedings of the Prussian Academy of Sciences, 1914. Reprint in *Ideas and Opinions*, based on Carl Seelig (Ed.), *Mein Weltbild*, and other sources. New translations and revisions by Sonja Bargmann, pp. 220–223. New York: Crown Publishers, 1954.

Fang, J. and Frønsdal, C. (1979), 'Deformations of gauge groups. Gravitation', *Journal of Mathematical Physics* **20**, 2264–2271. DOI: 10.1063/1.524007.

Feynman, R., Morinigo, F., Wagner, W. and Hatfield, B. (1995), *Feynman Lectures on Gravitation*, Addison-Wesley, Reading, Massachusetts.

Fierz, M. and Pauli, W. (1939), 'Relativistic wave equations for particles of arbitrary spin in an electromagentic field', *Proceedings of the Royal Society of London A* **173**, 211–232. DOI: 10.1098/rspa.1939.0140.

Friedman, M. (1983), *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*, Princeton University Press, Princeton.

Ghins, M. and Budden, T. (2001), 'The principle of equivalence', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **32**(1), 33–51. DOI: 10.1016/S1355-2198(00)00038-1.

Gupta, S. N. (1952*a*), 'Quantization of Einstein's gravitational field: General treatment', *Proceedings of the Physical Society. Section A* **65**(8), 608–619. DOI: 10.1088/0370-1298/65/8/304.

Gupta, S. N. (1952*b*), 'Quantization of Einstein's gravitational field: Linear approximation', *Proceedings of the Physical Society. Section A* **65**(3), 161–169. DOI: 10.1088/0370-1298/65/3/301.

Gupta, S. N. (1954), 'Gravitation and electromagnetism', *Physical Review* **96**(6), 1683–1685. DOI: 10.1103/PhysRev.96.1683.

Gupta, S. N. (1957), 'Einstein's and other theories of gravitation', *Reviews of Modern Physics* **29**(3), 334–336. DOI: 10.1103/RevModPhys.29.334.

Knox, E. (2019), 'Physical relativity from a functionalist perspective', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **67**, 118–124. DOI: 10.1016/j.shpsb.2017.09.008.

Kraichnan, R. H. (1955), 'Special relativistic derivation of generally covariant gravitation theory', *Physical Review* **98**(4), 1118–1122. DOI: 10.1103/PhysRev.98.1118.

Kraichnan, R. H. (1956), 'Possibility of unequal gravitational and inertial masses', *Physical Review* **101**(1), 482–488. DOI: 10.1103/PhysRev.101.482.

Lehmkuhl, D. (2008), Is spacetime a gravitational field?, *in* D. Dieks, ed., 'Philosophy and Foundations of Physics Vol. 4: The Ontology of Spacetime Vol. II', Elsevier, Amsterdam, pp. 83–110.

Lehmkuhl, D. (2019), 'General relativity as a hybrid theory: The genesis of einstein's work on the problem of motion', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **67**, 176–190. DOI: 10.1016/j.shpsb.2017.09.006.

Lehmkuhl, D. (2020), 'The equivalence principle(s)'. Manuscript. Forthcoming in E. Knox and A. Wilson (Eds.), *The Routledge Companion to Philosophy of Physics*. London: Routledge.

Linnemann, N. and Salimkhani, K. (2020), 'The constructivist's programme and the problem of pregeometry'. Manuscript.

Maggiore, M. (2008), *Gravitational Waves. Volume 1: Theory and Experiment*, Oxford University Press, Oxford.

Malament, D. B. (2012), A remark about the "geodesic principle" in general relativity, *in* M. Frappier, D. H. Brown and R. DiSalle, eds, 'Analysis and Interpretation in the Exact Sciences. Essays in Honour of William Demopoulos', Springer, Dordrecht, pp. 245–252.

Martens, N. and Lehmkuhl, D. (2020*a*), 'Dark matter = modified gravity? scrutinising the spacetime-matter distinction through the modified gravity/ dark matter lens'. Manuscript.

Martens, N. and Lehmkuhl, D. (2020*b*), 'A roadmap to interpreting fields that are both (dark) matter and (aspects of) spacetime'. Manuscript.

Mattingly, D. (2005), 'Modern tests of Lorentz invariance', *Living Reviews in Relativity* **8**, 5. DOI: 10.12942/lrr-2005-5.

Maudlin, T. (2012), *Philosophy of Physics: Space and time*, Princeton University Press, Princeton.

Menon, T. (2019), 'Algebraic fields and the dynamical approach to physical geometry', *Philosophy of Science* **86**(5), 1273–1283. DOI: 10.1086/705508.

Misner, C. W., Thorne, K. S. and Wheeler, J. A. (1973), *Gravitation*, Princeton University Press, Princeton.

Morrison, M. (2006), 'Emergence, reduction, and theoretical principles: Rethinking fundamentalism', *Philosophy of Science* **73**(5), 876–887. DOI: 10.1086/518746.

Nicolis, A. (2011), 'General relativity from Lorentz invariance'. Lecture notes.

North, J. (2018), A new approach to the relational–substantival debate, *in* K. Bennett and D. W. Zimmerman, eds, 'Oxford Studies in Metaphysics', Vol. 11, Oxford University Press, Oxford, pp. 3–43.

Norton, J. D. (2008), 'Why constructive relativity fails', *The British Journal for the Philosophy of Science* **59**(4), 821–834. DOI: 10.1093/bjps/axn046.

Ogievetsky, V. I. and Polubarinov, I. V. (1965), 'Interacting field of spin-2 and the Einstein equations', *Annals of Physics* **35**(2), 167–208. DOI: 10.1016/0003-4916(65)90077-1.

Ortín, T. (2015), *Gravity and Strings*, Cambridge University Press, Cambridge.

Ortín, T. (2017), 'Higher order gravities and the strong equivalence principle', *Journal of High Energy Physics* **152**. DOI: 10.1007/JHEP09(2017)152, arXiv:1705.03495v2 [gr-qc].

Padmanabhan, T. (2008), 'From gravitons to gravity: Myths and reality', *International Journal of Modern Physics D* **17**(03n04), 367–398. DOI: 10.1142/S0218271808012085.

Papapetrou, A. (1948), 'Einstein's theory of gravitation and flat space', *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences* **52**, 11–23.

Pauli, W. (1941), 'Relativistic field theories of elementary particles', *Reviews of Modern Physics* **13**(3), 203–232. DOI: 10.1103/RevModPhys.13.203.

Penrose, R. (1980), On Schwarzschild causality—a problem for "Lorentz covariant" general relativity, *in* F. J. Tipler, ed., 'Essays in General Relativity—A Festschrift for Abraham Taub', Academic, New York, pp. 1–12.

Pitts, J. B. (2016), 'Einstein's equations for spin 2 mass 0 from Noether's converse Hilbertian assertion', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **56**, 60–69. DOI: 10.1016/j.shpsb.2016.08.008.

Pitts, J. B. and Schieve, W. C. (2001*a*), 'Light cone consistency in bimetric general relativity', *AIP Conference Proceedings* **586**(1), 763–765. DOI: 10.1063/1.1419652.

Pitts, J. B. and Schieve, W. C. (2001*b*), 'Null cones in Lorentz-covariant general relativity', arXiv: 0111004 [gr-qc].

Pitts, J. B. and Schieve, W. C. (2001*c*), 'Slightly bimetric gravitation', *General Relativity and Gravitation* **33**, 1319–1350. DOI: 10.1023/A:1012005508094.

Pitts, J. B. and Schieve, W. C. (2007), 'Universally coupled massive gravity', *Theoretical and Mathematical Physics* **151**, 700–717. DOI: 10.1007/s11232-007-0055-7.

Poisson, E. and Will, C. (2014), *Gravity. Newtonian, Post-Newtonian, Relativistic*, Cambridge University Press, Cambridge.

Pooley, O. (2013), Substantivalist and relationalist approaches to spacetime, *in* R. Batterman, ed., 'The Oxford Handbook of Philosophy of Physics', Oxford University Press, Oxford.

Pooley, O. (2017), Background independence, diffeomorphism invariance, and the meaning of coordinates, *in* D. Lehmkuhl, G. Schiemann and E. Scholz, eds, 'Towards a theory of spacetime theories', Birkhäuser, Basel, pp. 105–143.

Preskill, J. and Thorne, K. S. (1995), Foreword, *in* R. P. Feynman, F. B. Morinigo, W. G. Wagner and B. Hatfield, eds, 'Feynman Lectures on Gravitation', Addison-Wesley, Reading, Massachusetts, pp. vii–xxx.

Read, J. (2016), '*Background Independence in Classical and Quantum Gravity*'. Master's Thesis, University of Oxford. URL: `https://ora.ox.ac.uk/objects/uuid:b22844af-aac9-4adc-a6c5-1e2815c59655`.

Read, J. (2018), 'Explanation, geometry, and conspiracy in relativity theory'. Forthcoming in C. Beisbart, T. Sauer and C. Wüthrich (Eds.), *Thinking About Space and Time: 100 Years of Applying and Interpreting General Relativity*, Einstein Studies series. Basel: Birkhäuser, forthcoming. PhilSci-Archive: 15253.

Read, J. (2019), 'On miracles and spacetime', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **65**, 103–111. DOI: 10.1016/j.shpsb.2018.10.002.

Read, J., Brown, H. R. and Lehmkuhl, D. (2018), 'Two miracles of general relativity', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **64**, 14–25. DOI: 10.1016/j.shpsb.2018.03.001.

Redhead, M. (1999), Quantum field theory and the philosopher, *in* T. Y. Cao, ed., 'Conceptual Foundations of Quantum Field Theory', Cambridge University Press, Cambridge, pp. 34–40.

Rosen, N. (1940*a*), 'General relativity and flat space. I', *Physical Review* **57**(2), 147–150. DOI: 10.1103/PhysRev.57.147.

Rosen, N. (1940*b*), 'General relativity and flat space. II', *Physical Review* **57**(2), 150–153. DOI: 10.1103/PhysRev.57.150.

Salimkhani, K. (2018), Quantum gravity: A dogma of unification?, *in* A. Christian, D. Hommen, N. Retzlaff and G. Schurz, eds, 'Philosophy of Science: Between the Natural Sciences, the Social Sciences, and the Humanities', Springer International Publishing, Cham, pp. 23–41. DOI: 10.1007/978-3-319-72577-2_2.

Salimkhani, K. (2019), 'Explaining unification in physics internally', *Synthese* . DOI: 10.1007/s11229-019-02436-x.

Schiff, L. I. (1960), 'On experimental tests of the general theory of relativity', *American Journal of Physics* **28**(4), 340–343. DOI: 10.1119/1.1935800.

Sexl, R. U. and Urbantke, H. K. (2002), *Gravitation und Kosmologie*, Spektrum, Heidelberg.

Stevens, S. (2018), 'Regularity relationalism and the constructivist project', *The British Journal for the Philosophy of Science* **71**(1), 353–372. DOI: 10.1093/bjps/axx037.

Thirring, W. E. (1959), 'Lorentz-invariante Gravitationstheorien', *Fortschritte der Physik* **7**(2), 79–101. DOI: 10.1002/prop.19590070202.

Thirring, W. E. (1961), 'An alternative approach to the theory of gravitation', *Annals of Physics* **16**(1), 96–117. DOI: 10.1016/0003-4916(61)90182-8.

Wald, R. M. (1986), 'Spin-two fields and general covariance', *Physical Review D* **33**, 3613–3625. DOI: 10.1103/PhysRevD.33.3613.

Weinberg, S. (1964*a*), 'Derivation of gauge invariance and the equivalence principle from Lorentz invariance of the S-matrix', *Physics Letters* **9**(4), 357–359. DOI: 10.1016/0031-9163(64)90396-8.

Weinberg, S. (1964*b*), 'Photons and gravitons in S-matrix theory: Derivation of charge conservation and equality of gravitational and inertial mass', *Physics Review* **135**(4B), B1049–B1056. DOI: 10.1103/PhysRev.135.B1049.

Weinberg, S. (1965*a*), 'Photons and gravitons in perturbation theory: Derivation of maxwell's and Einstein's equations', *Physics Review* **138**(4B), B98–B1002. DOI: 10.1103/PhysRev.138.B988.

Weinberg, S. (1965*b*), The quantum theory of massless particles, *in* S. Deser and K. Ford, eds, 'Lectures on Particles and Field Theory', Brandeis Summer Institute in Theoretical Physics, Prentice-Hall, Englewood Cliffs, New Jersey, pp. 405–485.

Weinberg, S. (1972), *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, Wiley, New York.

Weinberg, S. (1995), *The Quantum Theory of Fields. Volume I: Foundations*, Cambridge University Press, Cambridge.

Weinberg, S. (1999), What is quantum field theory, and what did we think it was?, *in* T. Y. Cao, ed., 'Conceptual Foundations of Quantum Field Theory', Cambridge University Press, Cambridge, pp. 241–251.

Will, C. (2001), *Theory and Experiment in Gravitational Physics*, Cambridge University Press, Cambridge.

Williams, P. (2017), 'Scientific realism made effective', *The British Journal for the Philosophy of Science* **70**(1), 209–237. DOI: 10.1093/bjps/axx043.

Wyss, W. (1965), 'Zur Unizität der Gravitationstheorie', *Helvetica Physica Acta* **38**(5), 469–480. DOI: 10.5169/seals-113605.