How to Teach General Relativity*

Guy Hetzroni† and James Read‡

Abstract

Supposing that one is already familiar with special relativistic physics, what constitutes the best route via which to arrive at the architecture of the general theory of relativity? Although the later Einstein would stress the significance of mathematical and theoretical principles in answering this question, in this article we follow the lead of the earlier Einstein (*circa* 1916) and stress instead how one can go a long way to arriving at the general theory via inductive and empirical principles, without invoking presumptions concerning the geometrical structure of the final theory. We focus on the construction of the kinematical structure and the terms describing the coupling of matter to gravity. General covariance, understood and employed as a straightforward extrapolation of empirical considerations, is central to our derivation, together with what we dub the 'Methodological Equivalence Principle'. We argue that our approach has a number of virtues, both for one's understanding of the general theory of relativity, but also for pedagogy, since it stresses—to the greatest extent possible (a lesson which we inherit from Bell [1976])—both the methodological precedence of dynamical considerations to interpretative issues and the theoretical continuity between general relativity and its precursors. We conclude by comparing our approach to other philosophical approaches to general relativity and discussing the significance of empirically motivated methodological principles in the philosophy of science.

Contents

*This is a preprint version of an article accepted for publication in *British Journal for the Philosophy of Science*.
†Department of Natural Sciences, The Open University of Israel, Israel. guyhe@openu.ac.il
‡Faculty of Philosophy, University of Oxford, United Kingdom. james.read@philosophy.ox.ac.uk

1 Introduction

Einstein's construction of the general theory of relativity is arguably the paradigm case of successful theorizing in physics: a process whereby persistent application of theoretical considerations with minimal empirical input leads to an empirically successful and mathematically unambiguous completed theory, the structure of which should be regarded as being exemplary for future theorising. However, almost every single aspect of the methodological and theoretical foundations of the theory soon became a matter of serious and ongoing controversy: the controversial issues included general covariance [Norton, 1993], the principle of equivalence [Norton, 1985], the status of geometry [Lehmkuhl, 2014], and the role of Mach's principle [Barbour and Pfister, 1995]. Nevertheless, what appears to be far less controversial is that general relativity brought a methodological shift into theoretical physics, pushing it away from the realm of experience and into that of mathematics and abstract theorizing. This shift includes an emphasis on (i) mathematically-based patterns of reasoning, (ii) construction of mathematical structures shaped by conceptions of simplicity and naturalness, and (iii) non-empirical guiding principles (e.g. general covariance) in the context of discovery. This shift is often associated with a philosophical acknowledgement of the role of *a priori* reasoning in science [Friedman, 1983, Ryckman, 2005]. The early debates on this issue between positivist, neo-Kantian, and conventionalist approaches shaped in a fundamental way the philosophy of science which was to follow.

In his later reflections, Einstein embraced and promoted the view of there being a radical shift in methodology which occured during the development of general relativity. In his Oxford talk of 1933, for instance, he argued that

> [t]he scientists [of the eighteenth and nineteenth centuries] were for the most part convinced that the basic concepts and laws of physics were [...] derivable by abstraction, i.e. by a logical process, from experiments. It was the general theory of relativity which showed in a convincing manner the incorrectness of this view. [...] The axiomatic basis of theoretical physics cannot be an inference from experience, but must be free invention [...] [E]xperience of course remains the sole criterion of the serviceability of a mathematical construction for physics, but the truly creative principle resides in mathematics. [Einstein, 1934, pp. 166–167]

And a year later,

> [t]he theory of relativity is a fine example of the fundamental character of the modern development of theoretical science. The initial hypotheses become steadily more abstract and remote from experience. [...] The predominantly inductive methods appropriate to the youth of science are giving place to tentative deduction. [Einstein, 1954a, p. 282].

By 'inductive methods', Einstein means the development of physical theories on the basis of extrapolation from observed empirical regularities—his 1905 formulation of special relativity, identified later by Einstein [1919] as a 'principle theory', would fit this mould. Indeed, the later Einstein was consistent in his devaluation of inductive reasoning in favor of mathematical forms of reasoning, and applied this view in his research programs. Arguably for these reasons, he misrepresented his earlier derivations of general relativity [van Dongen, 2010], which he described in his [1919] as based on "empirically discovered" elements formulated as principles. In contrast to other views presented by Einstein, including his portrayal of the individual principles of general relativity as well as the later ideas underlying his unified field program, the view presented in the text above is in harmony with the views and practice of other prominent contemporary physicists—for example, Weyl, Eddington, and Dirac—and, furthermore, would become increasingly dominant in the development of later physical theories.

The shift in the theoretical physics is often associated with a corresponding shift in the philosophy of science. The appeal to the *a priori* on the one hand, and on the other hand the gap between the relevant *a priori* notions and the ones that were familiar from the Kantian program, presented an epistemic challenge that had a fundamental influence on the philosophy of science from the 1920s onwards. Accordingly, the different schools offered competing views on the epistemological foundations of the theory. The neo-Kantian program appealed to the newly-introduced notion of the relativized *a priori*, starting with early Reichenbach [1921/1965] and Cassirer [1922], leading in turn to more recent accounts of advance and knowledge in physics [Ryckman, 2005, Friedman, 1983]. A different approach was taken by later Reichenbach [1924/1969], who aimed to suggest a more rigorous version of Einstein's construction, which, in light of logical positivism, replaced the principles with axioms allegedly extrapolated from 'elementary facts', supported by coordinative (operational) definitions. A more rigorous approach by Ehlers et al. [2012] applied operational definitions for the various geometrical structures of general relativity, based on trajectories of freely falling bodies and light rays.

In this paper we highlight a path to general relativity that draws as much as possible from empirical considerations and from the already established special relativity, thus providing an alternative to the view that the theory necessarily calls for a radical break in the methods of physics. Our approach does not fall under any of the philosophical approaches mentioned above. Instead of aiming to connect the concepts of the new theory directly to either experience, fundamental principles, or hypotheses, it aims to derive and construct as much as possible from familiar theories (special relativity in this case). In offering a way into general relativity that maintains continuity with later as well as earlier theories and emphasizes the role of experience, we aim to reflect on a methodological question which precedes the other epistemological and interpretative issues debated in the context of that theory—but one which can nevertheless shed light on them.

Claims for a radical conceptual break arose already in the case of the special theory of relativity. John Bell [1976] complained in the opening paragraph of his famous paper on the theory, "[u]sually it is the discontinuity which is stressed, the radical break with more primitive notions of space and time". Bell suggested an alternative reconstruction of the theory, aiming to "emphasize the continuity with earlier ideas". While today Bell's paper is often identified with a dynamical (rather than geometrical) approach to special relativity, Bell's paper in fact does not mention or criticize a geometrical (or 'Minkowskian') reading of the theory. Indeed, Bell distances his claim from interpretational or ontological issues by emphasizing that "the facts of physics do not oblige us to accept one philosophy rather than the other. And we need not accept Lorentz's philosophy to accept a Lorentzian pedagogy." Bell's point, however, goes beyond the pedagogical message of his title 'How to teach special relativity' (hence our homage in the title of the current article); his paper can naturally be read as an essay on forms of reasoning and theory construction. Bell identifies the radical break view with Einstein's famous style of deriving the theory from hypotheses (more commonly referred to as postulates or axioms). This reasoning stands in contrast to the Lorentzian style of the approach presented by Bell, in which the lion's share of the road to the theory is straightforwardly paved by familiar preceding theories. While our message does not depend in any way on Bell's view of special relativity,[1] we follow him in regarding theoretical and methodological continuity as a key to conceptual clarity. Moreover, we regard the project of revealing the particular ways in which a theory is continuous with earlier (and later) theories as an important part of the foundational understanding of the theory, without which the discussion on the relation between the concepts of a theory and

[1]The tenability of Bell's approach to special relativity can be questioned, e.g. due to Bradley's [2021] claim that Lorentz's approach does not provide a way to Einstein's special relativity but rather to a distinct theory. Since the relation to our claims is merely that of an analogy, such claims do not directly impact the message of the current paper. Note also that in contrast to Bell, our aim is not to dispel misunderstanding of the theory itself such as the one demonstrated by his spaceship example, but rather to reflect on the methodological lesson as in the rest of his paper.

its empirical content, or between a theory and the axioms or hypotheses that give rise to it, is essentially incomplete.

While it is debatable whether the introduction of general relativity involves a similar discontinuity in the concepts of space and time, when it comes to the methods of theoretical physics the break appears far more radical than that in special relativity. Riemannian geometry cannot simply be derived from experience, it seems to many—rather, it must either be postulated or derived from postulates that are more far-reaching than those of the special theory. General covariance, presented by Einstein as a fundamental principle of the theory, similarly appears as an hypothesis that can be motivated neither by experience nor by pre-relativistic theories. From a later perspective, though, geometry and invariance are typically understood as lying at the heart of the methodological shift associated with general relativity [see, e.g., Yang, 1980]. Thus, elucidating the methodological issues concerning the principles of the theory is essential not only for setting it on secure conceptual foundations, but also for understanding the interface of formal and empirical considerations in modern physics.

In this paper we show that special relativity, applied together with empirical considerations, can take us a significant part of the way towards relativistic gravity. We then further reflect on the theoretical considerations which are required to complete the derivation. These considerations, we argue, function as empirically-guided methodological principles, rather than fundamental hypotheses or axioms. This 'heuristics-first' approach permits the construction of the basic kinematical structure of general relativity together with the terms describing its coupling to matter, without making puzzling presuppositions regarding the mathematical form of the theory (and in particular on Riemannian geometry).

Close in several aspects to Einstein's [1916] early presentation of general relativity, the presented construction places the theoretical burden on the requirement of general covariance applied together with a modified version of the equivalence principle, both motivated and applied as methodological principles. Compared to some formulations of Einstein's equivalence principle, our 'Methodological Equivalence Principle' is presented as an empirically-motivated theoretical conjecture rather than a direct empirical observation. Its role is to prescribe a certain way of extending the empirical content of a theory in light of an invariance requirement supported by local evidence. When applied in the context of general covariance, the principle yields general relativistic coupling of matter to gravity, including a metric field and a derived field of the form of the Levi-Civita connection. This reconstruction turns out to be in close analogy with the introduction of coupling terms to free field theories using the gauge argument.

The structure of the article is this. In §2, we present briefly the conceptual debates regarding the equivalence principle, general covariance, and the role of geometry. In §3, we aim to examine how much of the general relativistic structure can be recovered by applying empirical considerations combining special relativity and Einstein's equivalence principle. In §4, we show how the approach can be naturally extended as a construction of a physical explanation of local inertial structure, and construct the geodesic equation. In §5, we generalize the approach into what we call the 'Methodological Equivalence Principle', show how it leads to general relativistic coupling terms, and discuss additional considerations that can lead to the field equations. §6 discusses the analogy between this introduction of gravitational coupling and the gauge argument. Finally, we conclude in §7 by revisiting the debates on general covariance, the principle of equivalence, and the role of geometry in general relativity, reflecting on the methodological lesson from the theory, and argue that empirical considerations supported by minimal theoretical considerations related to dynamics and representation provide a simple path to the theory in a way that is more inductive (in Einstein's sense) and tied to experience than is usually thought.

2 Equivalence, covariance, geometry, and dynamics

In the first part of Einstein's 1916 first review of general relativity, Einstein presents the fundamental ideas of the theory, which lead to the conclusion that the components of the metric tensor have "to be regarded from the physical standpoint as the quantities which describe the gravitational field in relation to the chosen system of reference" (p. 120). Einstein motivates and justifies this conclusion by appealing to the empirical equivalence of acceleration and gravity (that is, one version of his equivalence principle) together with the requirements of general covariance and the local validity of special relativity, alongside geometrical considerations. Einstein declared that the goal was not to present a rigorous derivation, but rather to develop the theory in a "psychologically natural" way (p. 118). The way in which these considerations are related to each other and to the conclusion was thus not stated explicitly. In this section we aim to review briefly relevant aspects of the foundational discourse on these considerations and the relation between them.

The development of general relativity started with the observation that "[a]s far as we know, the physical laws with respect to [a uniformly accelerating system] Σ_1 do not differ from those with respect to [a system at rest in a homogeneous gravitational field] Σ_2" [Einstein, 1907, p. 302].[2] Einstein's central motivation here was theoretical, namely to extend the principle of relativity to frames that accelerate with respect to the inertial frames of the special theory. Yet, the observation in itself is manifestly empirical, and would later be described as an observation promoted into a fundamental principle, 'Einstein's equivalence principle' [Einstein, 1919].

It is important to register that there are a number of different readings of what is being proposed regarding the connections between gravity and inertia. One such reading is that gravity is reduced to inertial effects. According to this approach, gravity is merely a pseudo-force characterizing non-inertial frames. A second option is that inertia is reduced to gravity. This possibility can be expressed as based on the notion of a gravitational field. The term 'gravitational field' can be identified with one of the mathematical concepts of the theory such as the metric, connection, or Riemman tensor [Lehmkuhl, 2008], or, as in Misner et al. [1973, p. 399], to "refer in a vague, collective sort of way to all of these entities". A third option is an "egalitarian view" according to which gravity and inertia are unified symmetrically without reducing one to the other, similar to the unification of electric and magnetic fields in special relativity. This option may best capture Einstein's own views in the period that followed his presentation of general relativity [Lehmkuhl, 2008, 2014, 2021].

The nature of the relation between inertia and gravity therefore characterises different interpretational approaches towards general relativity. Before elaborating further on this, however, let us note that in any case, this equivalence is far from sufficient to provide the basis for a new gravitational theory, as Einstein's struggles from 1907 to the completion of the theory in 1915 attest. Uniform acceleration does not have sufficient structure to account for the richness of gravitational effects. A stronger desideratum that appeared separately in the 1916 review is the local validity of special relativity, soon to be known as the 'strong equivalence principle'. The very validity of the principle in general relativity became controversial immediately after the publication of the theory, and Einstein reformulated it several times over the course of these dialogues [Norton, 1985].[3]

What seems to be less controversial, however, is that the equivalence principle, in one version or another, does have an heuristic value. For Einstein, the primary issue was an extension of the principle of relativity to accelerating frames. This can be seen in his later

[2] This observation was presaged in Newton's Corollary VI: for discussion, see Saunders [2013].

[3] In the modern debate, the guiding question is whether it is possible to formulate a version of the strong equivalence principle that would characterize general relativity in a non-trivial manner and from among competing theories, e.g. those that manifest torsion [Knox, 2013].

conceptual identification of gravitational effects and the effects of uniform accelerations (for discussion, see [Lehmkuhl, 2021]), so that one would thereby have constructed a candidate theory of gravitation necessarily invariant under all such transformations. Einstein sometimes presented these transformations as a special case of coordinate transformations, such that general covariance is a generalization of this idea.

Yet, the generalization from uniformly accelerating frames to arbitrary frames immediately became the subject of a similar controversy together with Einstein's conception that general covariance is a generalization of Lorentz invariance, and should be regarded as a relativity principle. The early debates on general covariance, beginning with Kretschmann [1917], were mainly focused on these claims [Norton, 1995]. Unlike Lorentz transformations, inertial effects render the transition to accelerating frames detectable. Furthermore, these transformations do not leave invariant the components of (say) the special relativistic Minkowski metric. These disanalogies make general covariance inherently different from familiar relativity principles. The generality of general covariance makes it, according to this criticism, devoid of physical content: a property of a formulation of a theory, not of its physical content. This seems particularly salient from the geometrical point of view, in which general covariance is a trivial outcome of expressing a theory using local coordinate-free variables of differential geometry. General covariance was useful for Einstein, according to Norton [1993], simply because these mathematical tools were unavailable for him at the time. Modern responses to these objections [e.g. Belot, 2011, Earman, 2006a,b, Pitts, 2006, Pooley, 2010, 2015] aim to identify a substantive (as opposed to artificial) notion of general covariance as a property that characterizes general relativity but not, for example, generally covariant formulations of special relativity or Newtonian gravity (for a recent survey, see [Read, 2023b]).

These different views on Einstein's equivalence principle, the strong equivalence principle, and the principle of general covariance, are not independent from the issue of geometry, and more widely from that of an interpretation of general relativity. Understanding Einstein's equivalence principle in terms of reduction of gravity to inertia typically leads to a geometrical outlook, according to which the theory demotes gravity from a force into an inertial effect, a consequence of spacetime geometry. This view is almost as old as the theory itself, and constitutes a common textbook presentation of the theory. The view is one possible interpretive stance towards the theory.[4] It is often associated with a geometrical view of special relativity in which the chronogeometrical significance of the Minkowski metric inherently reflects spacetime geometry.

According to an alternative 'dynamical' approach to special and general relativity, the chronogeometric significance of the metric tensor is not given *a priori* as part of the theory, but is rather to be derived [Brown, 2005]. Thus, while in the geometrical approach the kinematics of the theory are formulated in geometrical terms (and one might, e.g., stipulate that a metric field appearing in the kinematical possibilities of the theory is to be regarded as 'geometry'), in the dynamical approach the kinematics (insofar as one thinks that there is a clean kinematics/dynamics distinction to begin with) correspond to different configurations of various fields—metric fields, material fields, etc.—not *ab initio* distinguished from one another. Geometrical notions then have to be derived from the dynamics, taking the role of explanandum rather than explanans. Thus, only by dint of dynamical interactions that determine the readings of, say, rods and clocks does a metric field appearing in the kinematics inherit a geometrical interpretation [Huggett et al., 2022]. In general relativity, proponents of the dynamical view often ground the chronogeometric interpretation of the

[4]Einstein himself, however, took the opposite path, arguing that general relativity demonstrates that geometry, when applied to ordinary rigid objects, has to be understood as an empirical science rather than an infallible mathematical construct [Einstein, 1920, 1921]. Later, he criticized the geometrical view more explicitly, reducing the value of geometrical considerations to useful heuristics: geometrization, he claimed, "is only a kind of crutch for the finding of numerical laws" (cited by Lehmkuhl [2014], p. 317).

metric field in the strong equivalence principle, which goes beyond the content of the field equations of general relativity [see discusstion in Brown, 2005, §9.5.2].

What is of more relevance to our concerns is that the identification of general relativity with Riemannian geometry reflects on the construction of the theory as well as its understanding. Indeed, both tasks become much simpler if ones conjectures fron the start the applicability of Riemannian geometry! When the principle of general covariance and the principle of equivalence are criticized as trivial, dispensable, or lacking content, it is usually from this geometrical point of view. Synge [1960], for example, famously compared the equivalence principle to a midwife that should be "buried with appropriate honours" (p. x) now that her duty is done. Here, our primary concern is with the *construction* of general relativity, rather than with its interpretation or extension. Our aim is not to make an interpretative claim or to preclude a geometrical reading of general relativity, but rather to show that the theory can be constructed without making premature presuppositions about its geometrical or mathematical architecture. The possible contributions of our results to the dynamicist's project are discussed in §7.2.

It is also important to note at this point that there is no necessity to associate a geometrical (or dynamical) understanding of special relativity to a corresponding reading of the general theory. The relation between gravity and inertia (that determines the interpretation approach in general relativity) is not necessarily linked to the issue of the chronogeometrical significance of the Minkowski metric of special relativity. In particular, a dynamically-spirited reading of Einstein's equivalence principle (namely, any reading that avoids reducing gravity to inertia) is compatible with either a dynamical or a geometrical reading of the chronogeometrical significance of the Minkowski metric in special relativity.[5]

As we have seen, by combining the equivalence principle with general covariance, Einstein was able to make significant progress in arriving at a theory of gravity which held in *all* frames of reference. In other words, the above geometrical accounts may be insufficient to understand the methodological imprint of general relativity on theoretical physics.[6] This task requires an account of the heuristic significance of both general covariance and the equivalence principle. To develop such an account constitutes our central goal in the sections of this article to come.

3 Empirical paths towards general relativity

In a 1918 letter to Michelle Besso, Einstein tried to relax Besso's worries that the development of general relativity reveals that "speculation allegedly had revealed itself to be superior to empiricism." Einstein argues that in fact "no genuinely useful and profound theory has ever really been found purely speculatively." In the context of general relativity, he argues, the equivalence of inertia and gravity is the generalizable empirical fact at the basis of the theory. Indeed, among the different considerations discussed in the previous section, Einstein's equivalence principle is the only one that is clearly and directly linked to experience. Let us therefore examine the capacity and limits of promoting this observation into a theoretical principle, that of conceptual unification of gravity and inertia. Suppose that one embraces this principle. Suppose further that one has an antecedent commitment to special relativity, understood as the claim that all physics should be conditioned so as to be invariant under Lorentz boosts.[7] How far could one proceed with these two commitments

[5]A geometrical interpretation of the Minkowski metric could naturally be generalized into a geometrical reading of the metric tensor of general relativity (cf. footnote 18), but this is not a logical necessity. For example, one could regard the symmetries of non-gravitational laws as determined by local spacetime geometry while at the same time consider this local geometry as determined and explained by the gravitational field.

[6]Cf. Giovanelli [2021, pp. 4–5]

[7]One may hold such a commitment on the basis of (i) the discovery of electromagnetism and (ii) null results of experiments such as that of Michelson and Morley—i.e., on the basis of the reasons which Einstein adduced in 1905

alone?

As is well-known (see e.g. Brown and Read, 2016 for recent discussion), it follows from a commitment to the Einstein equivalence principle that rest frames of observers stationary on the surface of the Earth are not inertial frames of reference—i.e., frames of reference in which the laws of physics simplify maximally.[8] The reason for this is that, in the freely-falling frames (i.e., the frames co-moving with freely falling observers), it is no longer the case that gravitational effects are simply *cancelled* by inertial effects (this being the account given in theories such as Newtonian gravity).[9] Rather, in such frames there simply *are* no unified gravito-inertial effects (to take here Einstein's unified understanding of the equivalence principle discussed above), whereas in the rest frames of observers on the Earth's surface, there (of course) *are* gravitational effects. As a consequence, having embraced Einstein's equivalence principle, it is the freely falling frames which are the inertial frames.

So, in our hypothetical scenario, an observer situated on the surface of the Earth believes that (a) causal structure is relativistic—i.e., that of Minkowski spacetime, and (b) they are situated in a uniformly accelerating frame. Combining these two commitments, our observer believes that their rest frame is a *Rindler frame*—these indeed being the uniformly accelerating frames of special relativity. For a proper acceleration α in the positive vertical direction (treated as the component to be denoted by ρ), the line element, written in Rindler coordinates, reads (here, $\kappa := \alpha/c^2$)

$$ds^2 = -\frac{1}{\kappa^2}d\tau^2 + \frac{1}{\kappa^4}d\rho^2 + dy^2 + dz^2; \tag{1}$$

by contrast, for a freely falling observer, the line element will take the familiar form

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2. \tag{2}$$

(At this point, in the manner of [Einstein, 1905], one can maintain an operational understanding of the coordinates in both of these frames—i.e., one can maintain that coordinate intervals correspond to the readings of idealised rigid rods and clocks in each frame.) In Rindler coordinates, the components ${\Gamma^\mu}_{\nu\sigma}$ of the Levi-Civita connection do not vanish; rather, they take the form[10] [Müller and Grave, 2014, p. 17]

$${\Gamma^\rho}_{\tau\tau} = -\kappa, \qquad {\Gamma^\tau}_{\tau\rho} = -\frac{1}{\kappa}, \qquad {\Gamma^\rho}_{\rho\rho} = -\frac{2}{\kappa}; \tag{3}$$

all other components vanish.

Note that while above we have used the familiar geometric language of the line element and the connection, in fact all we are doing here is special relativity in an accelerating frame (Einstein equivalence principle aside). A parallel path can be taken in a more dynamical spirit by setting off from the equations of motion of a free particle. Defining the velocity vector of a test body as $u^\mu := \frac{dx^\mu}{d\lambda}$, where λ is the proper time parameter along the particle's worldline, in freely-falling coordinates the equations of motion are

$$\frac{du^\mu}{d\lambda} = 0, \tag{4}$$

whereas in Rindler coordinates we have

$$\frac{du^\mu}{d\lambda} - {\Gamma^\mu}_{\alpha\beta}u^\alpha u^\beta = 0, \tag{5}$$

for special relativity understood in this manner in the actual world.

[8]For more on the definition of inertial frames of reference, see Knox [2013].

[9]Here, we idealise to a uniform gravitational field—i.e, ignore tidal effects in the gravitational field of the Earth.

[10]Below, τ, ρ do not denote free indices, but rather denote the τ-component (i.e., 0-component) and ρ-component (i.e., 1-component) in the salient index.

where connection coefficients are as in (3). Of course, at this point, we have merely used special relativity to rewrite our physics in an accelerating frame—though *nota bene*: doing so is still sufficient to explain certain effects (e.g., the results of gravitational redshift Pound-Rebka experiments) without appeal to more sophisticated notions such as spacetime curvature: on this, see [Brown and Read, 2016].

How to move beyond the case of a uniform gravitational field? Consider now another observer, situated at some other point on the surface of the Earth. By parity of reasoning, their rest frame is also (given the Einstein equivalence principle) a non-inertial frame, so they can explain observed effects (equivalently) in terms of the influence of a gravitational field. But now, since we know (let us assume on the basis of other experiments and operational procedures) that these observers are situated at different points on the Earth's surface, we know that said gravitational field must be non-uniform, and accordingly, inertial structure is local. In other words, one would generalise the coefficients in (3) to be functions of space and time. Note that we have arrived at this conclusion using solely special relativity and the Einstein equivalence principle; moreover, note that one could arrive at this result without any background knowledge of (say) Newtonian gravitation theory.[11]

4 Empirical considerations meet mathematical representations

The empirical considerations presented in the previous section take us a step forward in the journey from special relativity into the general theory. We move now to ask: what minimal theoretical considerations must be added to these empirical considerations in order to proceed further towards general relativity?

Let us examine the above considerations in more detail. (5) can be understood as manifesting any of the possible relations between inertia and gravity presented in §2. If Einstein's equivalence principle is taken to imply the reduction of gravity to inertia, then the step taken in the previous section can be regarded as transforming the law of inertia from an inertial (freely falling) frame, into a non-inertial frame, to explain the outcomes of experiments in the observer's frame due to a 'gravitational' pseudo-force. The field interpretation would regard the connection coefficients as a manifestation of a gravitational field that creates deviation from pure inertial motion. In the egalitarian understanding, the two terms stand for a frame-dependent division of the unified phenomena into gravitational and inertial. In other words, different vertically-accelerating observers describe the same local phenomena using (5) with different values of κ, thus manifesting different 'mixtures' of inertia and gravity in their description (cf. Lehmkuhl [2014], §4).

One way to understand the content of (5) (together with the corresponding transformation law between different accelerating frames) is therefore as a determination of the local inertial frames by a physical object (such as a field) whose frame dependent representation is given by the Γ coefficients. This understanding coheres well with the field and egalitarian approaches (although possibly not necessarily with the geometrical reading of the equivalence principle, according to which inertia is a primitive notion).[12] According to this understanding, (5) is not merely providing a description for observed phenomena in the given frame which happens to be occupied by human observers; rather, it also explains why certain observers are inertial rather then others from among those that correspond to different accelerating observers. This line of reasoning regards the preferred 'inertial' frames

[11] For further discussion of these issues, see [Fankhauser and Read, 2023].

[12] Note that while our suggestion to understand (5) as describing the local determination of inertial frames (and its extension below) seems potentially to be in tension with a geometrical reading of Einstein's equivalence principle, it does not necessarily conflict with a geometrical understanding of the Minkowski metric in special relativity.

as an explanandum, a semi-empirical concept that needs to be determined by the physical content of the theory.[13]

To take up this point, let us note that the role of frames can most broadly be understood in this context as references with respect to which coordinates are assigned to events in spacetime. Thus, explaining the special status of the inertial frames amounts to seeking an explanation as to why some coordinatizations of spacetime lead to simpler formulation of the laws than others. At this point, one clearly steps further beyond strict observational considerations. The concept of 'inertia' is conceived here not in terms of its empirical content, but rather in terms of its mathematical representation in special relativity. It is an explanandum that has one foot in the theoretical structure (inertial coordinates), and the other one in experience (telling us that the theory is successful, and also which particular frames are the inertial ones). The considerations presented in the previous section suggest that inertia is physically determined. Now we aim to show how it is determined as a property of certain coordinatizations, not of particular observers. Thus, while arbitrary coordinate systems are related to the requirement for general covariance, in our construction we move to the space of arbitrary coordinate systems only to identify within it the familiar inertial coordinatizations.

The mathematical concept that can describe the physical determination of the set of local inertial frames can be constructed in the following way. To illustrate, begin with the Lagrangian describing the motion of a test particle in an inertial frame of reference:[14]

$$\mathscr{L} = \eta_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} =: \eta_{\alpha\beta}u^\alpha u^\beta. \tag{6}$$

In arbitrary curvilinear coordinates, the same Lagrangian takes the form

$$\mathscr{L} = \omega_{\mu\nu}\frac{d\xi^\mu}{d\lambda}\frac{d\xi^\nu}{d\lambda}, \tag{7}$$

with

$$\omega_{\mu\nu} := \eta_{\alpha\beta}\frac{\partial x^\alpha}{\partial \xi^\mu}\frac{\partial x^\beta}{\partial \xi^\nu}. \tag{8}$$

This simple mathematical transformation makes it easy to see that a physical object that defines a local set of inertial frames can be described by a field $g_{\mu\nu}$ for which the values of the components transform tensorially under coordinate transformation (namely, according to the same transformation law that turns $\eta_{\mu\nu}$ into $\omega_{\mu\nu}$), and its possible local values are the possible values of $\omega_{\mu\nu}$ defined in (8). Once $\eta_{\mu\nu}$ is replaced with this new field $g_{\mu\nu}$, the Lagrangian becomes

$$\mathscr{L} = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}. \tag{9}$$

This form guarantees that in any spacetime region which is sufficiently small (such that changes in $g_{\mu\nu}$ can be neglected), it would be possible to assign local coordinates x^μ in which the dynamics of free particles would be described by (6).

Notably, the field $g_{\mu\nu}$ has basic mathematical properties necessary to be interpreted as a metric (in particular, the requirement that there exist a coordinate transformation $x^\alpha \to \xi^\mu$ such that locally $g_{\mu\nu} = \omega_{\mu\nu} := \eta_{\alpha\beta}\frac{\partial x^\alpha}{\partial \xi^\mu}\frac{\partial x^\beta}{\partial \xi^\nu}$ together with Sylvester's law of inertia guarantee the correct signature).

[13] A reasonable worry at this point is that the notion of an inertial frame is not given to us in any straightforward unambiguous way, instead it is theoretically laden; moreover, what counts as an inertial frame might be fixed *a priori*, or as a matter of convention. This is a well-debated aspect of the relation between special and general relativity. We revisit these worries in §7.1.

[14] The reasons for our choice of Lagrangian formalism will be clarified further in §6.

The replacement of $\eta_{\mu\nu}$ with $g_{\mu\nu}$ has thus achieved three goals: first, the theory is now generally covariant. Second, it now has additional physical content. Third, the new physical content explains the non-invariance of the original local theory given by (6).

The equation of motion for $u^\mu := dx^\mu/d\lambda$ is the usual geodesic equation with the Levi-Civita connection. This follows from the Euler-Lagrange equation $\frac{d}{d\lambda}\frac{\partial\mathscr{L}}{\partial u^\mu} - \frac{\partial\mathscr{L}}{\partial x^\mu} = 0$ with the Lagrangian (9) via use of the chain rule:

$$\frac{d}{d\lambda}u^\sigma - \frac{1}{2}g^{\sigma\alpha}\left(\frac{\partial}{\partial x^\mu}g_{\alpha\nu} + \frac{\partial}{\partial x^\nu}g_{\nu\alpha} - \frac{\partial}{\partial x^\alpha}g_{\mu\nu}\right)u^\mu u^\nu =: \frac{d}{d\lambda}u^\sigma - \Gamma^\sigma_{\mu\alpha}u^\mu u^\nu$$
$$= 0. \qquad (10)$$

The procedure described above sets off from the aim to explain the existence of these preferred inertial frames. An important point has to be made regarding the locality of this explanation, which characterizes first and foremost the evidence. Local evidence for special relativity (e.g., experiements performed in particle accelerators) does not directly support the global validity of the theory with a universally fixed $\eta_{\mu\nu}$. This kind of evidence can just as well suggest that the explanation should be local. This notion of locality does not take the form of an *a priori* constraint and not even that of a theoretical virtue; it simply reflects the fact that the existing evidence supporting the preferred inertial frames is evidence collected locally. Even a single observer can make this observation (of course, two distant observers comparing their findings can note directly that the set of inertial frames in different spacetime points are not trivially related to each other). The introduction of degrees of freedom that determine inertial structure locally, as described above, is natural given the local nature of the evidence for the special theory. We turn now to showing that the considerations described here can be applied as a general theoretical method, a version of the equivalence principle.

5 The midwife resurrected; general covariance rehabilitated

Famously, reflecting on the development of the general theory of relativity, Einstein described the observation that led to the formulation of his equivalence principle as "the happiest thought of my life" [Einstein, 1920, p. 136]. This is an empirical observation: an equivalence between an observer in a free fall and an observer in zero gravity. Later in the paper, however, Einstein notes that the observation can be turned on its head and used as a theoretical principle:

> If we know the laws of nature with respect to a (gravitation-free) system K, then we can by mere transformation learn the laws relative to [a uniformly accelerated coordinate system] K', i.e., we learn about the physical properties of a gravitational field by means of a purely speculative method. (p. 137)

This is a remarkable speculation: the mathematical transformation of the representation of a physical situation from one coordinate system to another informs us about the properties of a physical field—one that was absent in the original situation! This speculative terminology can be epistemologically worrisome (cf. Einstein's use of the term in the letter to Besso quoted on the beginning of §3). These worries may be somewhat eased if the speculative method is understood as a suggested heuristic. In this section we aim to show how such an heuristic can be justified in light of the previous sections, and to spell it out in a clear and generalizable way. This attempt can be motivated beyond understanding general relativity: if Einstein's equivalence principle has any value for the purpose of unification it is in this speculative form, that does not depend on pre-existing knowledge of the interaction that originates in macroscopic phenomena (and has no parallel in the case of the nuclear forces, for example).

This way of applying the equivalence principle sets off from insisting on the equivalence of the descriptions of events given in two frames that uniformly accelerate with respect to each other. These events may involve non-gravitational forces. The appearance of pseudo-forces in the dynamical law that corresponds to certain frames seems to break the equivalence. The insistence on restoring it gives rise to the speculation of an additional force whose magnitude is proportional to the inertial mass. Thus, the equivalence of inertial and gravitational mass, and possibly the very existence of the gravitational interaction, is derived from the invariance requirement, rather than known in advance from experience. This argument would lead one to expect new phenomena, such as gravitational time dilation and bending of light-rays due to the gravitational field. (Needless to say, in this reasoning these phenomena remain speculations until observed.)

This sense of the equivalence principle goes beyond unifying gravity and inertia: the emphasis is theoretical, as a conjecture that expands the theory in a certain way that would guarantee invariance under the given transformation. This form of theoretical conjecture brings to mind the gauge argument, which similarly introduces an interaction based on an invariance requirement. Such an account of the gauge argument was given for example by Hetzroni [2021], with the aim of emphasizing the physical nature of the conjecture, above and beyond the mathematical notion of invariance. The account is described by Hetzroni [2020] as a generalized equivalence principle. According to this principle, newly-conjectured interaction terms are added in the place of terms that break the invariance of the dynamics under a change of mathematical representation. In this way, an invariant law describing the interaction is constructed based on the non-invariance of the interaction-free dynamics.[15] A formulation that works in the more abstract case of gauge symmetries may just as well work with coordinate transformations, and allow us to apply Einstein's "speculative method" in the case of general covariance. This generalized equivalence principle can be given a more careful formulation as a methodological principle in the following way:[16]

> *Methodological Equivalence Principle*: Given a non-invariant dynamical law in the sense that its form simplifies maximally in a given preferred class of representations but involves modified/additional expressions in arbitrary representations, construct an invariant law by replacing the modified/additional expressions with new dynamical fields, whose set of possible local values is identical to that of the modified/additional expressions, and which manifest the same representation-to-representation transformation properties.

Our aim now is to show that (1) this method can easily be generalized to the case of coordinate transformations, and that (2), this generalization is the exact manifestation of the simple considerations described in the previous section, concerning the generalization of local evidence about preferred representations.

Consider, then, an arbitrary curvilinear coordinate transformation applied to the Lagrangian of a Klein-Gordon scalar field:

$$\begin{aligned}\mathscr{L} &= \frac{1}{2}\left[\eta^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - m^2\phi^2\right] \\ &= \frac{1}{2}\left[\eta^{\alpha\beta}\left(\frac{\partial\xi^\mu}{\partial x^\alpha}\frac{\partial\phi}{\partial\xi^\mu}\right)\left(\frac{\partial\xi^\nu}{\partial x^\beta}\frac{\partial\phi}{\partial\xi^\nu}\right) - m^2\phi\right] \\ &= \frac{1}{2}\left[\omega^{\mu\nu}\frac{\partial\phi}{\partial\xi^\mu}\frac{\partial\phi}{\partial\xi^\nu} - m^2\phi\right],\end{aligned} \tag{11}$$

[15] See also Hetzroni and Stemeroff [2023] for a discussion of the analogy between the interactions.

[16] In reaction to this formulation of the Methodological Equivalence Principle, and following the lead of e.g. [Fletcher, 2020, Fletcher and Weatherall, 2023a,b, Weatherall, 2020], one might worry about the invocation of the notion of 'simplicity' of equations. We address this issue in §7.1.

with

$$\omega^{\mu\nu} := \eta^{\alpha\beta} \frac{\partial \xi^{\mu}}{\partial x^{\alpha}} \frac{\partial \xi^{\nu}}{\partial x^{\beta}}. \tag{12}$$

(Note that this $\omega^{\mu\nu}$ is the inverse of the $\omega_{\mu\nu}$ appearing in (8).) Applying the Methodological Equivalence Principle by now replacing $\omega^{\mu\nu}$ with $g^{\mu\nu}$, one obtains the Lagrangian

$$\mathscr{L} = \frac{1}{2}\left[g^{\alpha\beta}\partial_{\alpha}\phi\partial_{\beta}\phi - m^2\phi^2\right]. \tag{13}$$

Thus, one is once again led to a generally covariant action featuring a new field $g_{\mu\nu}$.

Up to this point, the $g_{\mu\nu}$ field introduced is a fixed object. In order to provide a genuinely physical explanation for the appearance of inertial frames, the field also has to be contingent—it should be conceived as a dynamical variable, one which is influenced by other fields and afforded its own dynamics (see below).[17]

Now, varying to obtain the Euler-Lagrange equation yields:

$$\begin{aligned} -m^2\phi &= \left(\partial_{\lambda} g^{\lambda\nu}\right)\frac{\partial\phi}{\partial x^{\nu}} + g^{\lambda\nu}\frac{\partial\phi}{\partial x^{\lambda}\partial x^{\nu}} \\ &= g^{\lambda\sigma}\Gamma^{\nu}{}_{\lambda\sigma}\frac{\partial\phi}{\partial x^{\nu}} + g^{\lambda\nu}\frac{\partial\phi}{\partial x^{\lambda}\partial x^{\nu}} \\ &= g^{\lambda\nu}\phi_{;\nu\lambda}. \end{aligned} \tag{14}$$

In the final two transitions we have used the definition of the Levi-Civita connection and the semicolon derivative to obtain the standard form of Klein-Gordon equation in curved spacetime.

In the construction of the geodesic equation in §4 and the above construction of the Klein-Gordon equation with general relativistic gravitational coupling, no assumptions have been made about the geometric structure of the resulting theory. The field $g_{\mu\nu}$ was introduced in this derivation as a field that restores invariance while at the same time explaining local inertial effects. Similarly, the coupling prescription at which we arrive does not presume the replacement of coordinate derivatives with covariant derivatives: the covariant derivative in (14) is *derived*, rather than presupposed. Once the $g_{\mu\nu}$ field and the coupling has been introduced into the theory, anyone familiar with Riemannian geometry can interpret the new concepts as a metric field and a covariant derivative featuring the Levi-Civita connection. The important point from an heuristic point of view is that it is not necessary to presuppose these geometrical notions in advance in order to construct these significant aspects of general relativity. Whether or not to adopt a geometrical understanding of the resulting theory is now a matter of various considerations not related to the indispensability of a geometrical perspective for the construction of the theory.[18]

The physical content of general covariance is therefore revealed not as a *formal* requirement, but rather as an *heuristic* one, which gains its significance only when applied together with the Methodological Equivalence Principle. This leads to a general relativistic kinematics (i.e., a $g_{\mu\nu}$ field with the correct signature and transformation properties), and also to the correct general relativistic coupling prescription.

Although the construction of the kinematic structure and coupling terms are the basis for our discussion in §6 and §7, it is worth dedicating a short side note to the possible ways in which to construct the field equations that would render the theory complete. There are many ways in which one might select a suitable dynamics, and indeed such dynamical

[17] Making this move can be motivated by further physical principles, for example Einstein's 'action-reaction principle' [on which see Brown and Lehmkuhl, 2016].

[18] That said, if one takes a geometrical stance towards the Minkowski metric of special relativity, and then proceeds to general relativity via the kinds of heuristics discussed in this article, then that might naturally in turn invite a geometrical stance towards the generalisation of the Minkowski metric so obtained—i.e., to the metric field $g_{\mu\nu}$ of general relativity. This, however, is not the only interpretational possibility; see footnote 5.

choices might be guided and constrained via other reasoning.[19] Such considerations are often implied at the level of the action. One example of such a path leading to the Einstein field equations is based on Lovelock's theorem [1971, 1972], which states that from a local action which contains at most second derivatives of $g_{\mu\nu}$, the only possible Euler-Lagrange equations of motion are the Einstein field equations. As for another example, one could take an effective field theory approach, considering all possible dynamical couplings in a Lagrangian describing local fields, and then identifying those terms relevant at a certain energy scale. Explicitly, one writes down an action of the form

$$S = \int d^4x\sqrt{g}\left(\frac{1}{16\pi G}R + c_1R^2 + c_2R_{\mu\nu}R^{\mu\nu} + \ldots + \mathscr{L}_{\text{matter}}\right), \tag{15}$$

before arguing that higher-order terms (i.e., those with coefficients $c_1, c_2, \ldots$) are irrelevant at low energies (see Donoghue [1994, 1995]). In this way, one can pick out the dynamical structure of general relativity (i.e., the Einstein equations) as the first-order result in an infinite energy expansion.[20] Note that in contrast to our introduction of the general-relativistic coupling terms, the above considerations do appeal to locality as a theoretical virtue. Relaxing this constraint can lead to additional nonlocal terms [e.g., Deser and Woodard, 2007]. Our suggested route into relativistic gravity therefore presents the theory as a chief possibility from among many possibilities, that is distinguished from them primarily based on certain theoretical virtues.

Does the Methodological Equivalence Principle deserve to be called an equivalence principle? In its straightforward minimalist formulation given above, it might seem like an unmotivated mathematical guess. In fact, however, the logic of the previous sections aims to show how it introduces exactly the field degrees of freedom that can identify dynamically a local class of preferred (i.e. inertial) coordinatizations from among all possible ones (more precisely, from among all those connected by a local coordinate transformation). In other words, the Methodological Equivalence Principle justifies its name as it underwrites the approximate validity of the strong equivalence principle; special relativity—the interaction-free theory in this case—is locally valid as long as the new $g_{\mu\nu}$ field is approximately constant in a region. When this is the case, one can identify a preferred class of coordinate systems (related by Poincaré transformations, in virtue of the spacetime symmetry of the interaction-free Lagrangian) in which physical laws recover their familiar special relativistic forms within that region. We revisit in §7 the issue of the relation between this principle and familiar formulations of the equivalence principle.

6 The heuristics of unification

The attempt to draw lessons from the success of general relativity was intertwined from the beginning with ambitions of unification. The most famous early attempts at unification consisted in classical 'unified field theories' such as the ones developed by Weyl [1918] (in that article, Weyl first introduced the terminology of 'gauge') and Einstein [1925, 1928a,b, 1931], alongside the Kaluza-Klein theory. Development of these ideas in a quantum context led, *inter alia*, to the first formulation of the modern gauge principle by Weyl [1929a,b], describing the coupling of spinors to electromagnetism and gravity using one principle presented using a geometric and group-theoretic emphasis.[21]

[19] For one famous catalogue of six different derivations of the dynamics of general relativity, see [Misner et al., 1973, ch. 17]; for a recent historical account of how Einstein himself found his field equations—a derivation which invoked both energy conservation and the Newtonian limit—see [Janssen and Renn, 2022].

[20] Allowed terms in (15) are constrained by invariance under what are regarded as being the relevant symmetries—in this case, diffeomorphisms. Note that in (15) we have omitted the cosmological constant term.

[21] See [O'Raifeartaigh, 1997] for the original papers and a detailed account of these developments.

The view that general relativity is the progenitor of a unification program based on formal and geometrical considerations is also a dominant theme in many later reflections. This view is most clearly expressed by Yang [1980], who identifies its starting point with the generalization from Lorentz transformations to general covariance. The relation of general covariance to gauge transformations, however, is a non-trivial matter that remains open—as does the more general question of the applicability of the concept of gauge to gravity (for more recent discussions see Wallace [2015], Weatherall [2016].)

In this section, we aim to show how the Methodological Equivalence Principle provides another angle on this issue, demonstrating that the gravitational coupling terms can be introduced into a gravitation-free special relativistic settings using heuristics that are analogous to those of the gauge argument. It is not the aim of this paper to reflect on whether and how this understanding can contribute to the unification program, but rather to present the analogy.

Just like general covariance, the requirement for the invariance of the Lagrangian under gauge transformations can be regarded by itself as constraining the formalism, not the physical content. Gauge invariance can be achieved without changing the content of a theory simply by replacing derivatives with gauge covariant derivatives that include a flat connection. The Methodological Equivalence Principle can easily be seen as the additional ingredient that turns the violation of the invariance requirement into a coupling prescription. Below, we present the analogy. We then turn to show that in both cases, the argument is based neither on mathematical necessity nor on purely mathematical guesswork. Instead, it is a generalization of local empirical evidence, and considerations that relate to the mathematical representation of the state and the dynamics.

Consider the standard gauge argument in the context of Yang-Mills theory. The starting point is two non-interacting spinor fields of equal mass described by the equations of motion

$$\sum_{a=1}^{2} \left(i\gamma^{\mu}\partial_{\mu}\psi_a - m\psi_a\right) = 0. \tag{16}$$

At this point the Yang-Mills field B^{μ} is commonly introduced in order to obtain covariance under local $SU(2)$ transformations. These transformations were presented by Yang and Mills [1954] as a change in the local convention which apply the labels '1' and '2' ("what to call a proton, what a neutron" in their terminology, p. 192) to describe the two fields. Let us examine this procedure in further detail. Our starting point is the observation that these equations of motion are valid in this form only in certain isospin conventions. An arbitrary change in the local isospin convention is represented by a coordinate dependent unitary matrix ${S_a}^b(x)$ that belongs to the relevant representation of the $SU(2)$ group, and can be expressed using Lie algebra: ${S_a}^b(x) = e^{-iT_j\beta_j}$. Under this transformation the numerical values representing the spinor fields transform according to $\psi_a \to \psi'_a = {(S^{-1})_a}^b\psi_b$ (spacetime dependence implicit). This transformation is not a symmetry due to the non-covariance of the derivative, which in the new isospin convention is written as

$$\begin{aligned}\partial_{\mu}\psi_a &= \partial_{\mu}\left({S_a}^b\psi'_b\right)\\ &= {S_a}^b\partial_{\mu}\psi'_b + (\partial_{\mu}{S_a}^b)\psi'_b\\ &= S\left(\partial_{\mu}\psi'_b - S^{-1}i(T_j\partial_{\mu}\beta_j)S\right)\psi'_b.\end{aligned} \tag{17}$$

Therefore, the form of the equations of motion in the new isopspin convention is

$$\sum_{a=1}^{2} i\gamma^{\mu}\left(\partial_{\mu}\psi'_b - iS^{-1}(T_j\partial_{\mu}\beta_j)S\psi'_b\right)\psi'_a - m\psi'_a = 0. \tag{18}$$

The standard gauge argument achieves invariance by introducing a compensating field B_{μ} whose transformation law yields a term that cancels out the one that emerges from the derivative. The original derivative and the compensating field are united to form the gauge covariant derivative

$$D_\mu \psi_b = \partial_\mu \psi_b - i\varepsilon B_\mu \psi_b. \tag{19}$$

The field can be written explicitly as ${B_{\mu a}}^b$, a connection over the $SU(2)$ principal bundle.

The Methodological Equivalence Principle applied in this context would lead to the same result, although in a slightly different way. According to the principle, we have to regard the covariance-violating term (i.e., the term $S^{-1}i(T_j\partial_\mu\beta_j)S$ in (17)) by itself as indicating the required change in the dynamical law. The way to achieve covariance is by introducing into the original equation a term $i\varepsilon {B_{\mu a}}^b\psi$ with the same index structure, set of possible local values, and transformation law.

The main difference between the standard gauge argument and an invariance imposed by applying the Methodological Equivalence Principle is that in the latter case we are not committed to the additive structure of the compensating field. §5 shows that the metric similarly restores invariance when it is introduced as a field that *multiplies* existing terms. In both cases, the new theory has additional empirical content for two reasons. First, the space of global field configurations is larger than the space of mathematical functions that are introduced to restore invariance: the possible values of the term involved in the mathematical transformation are the local values of the introduced field, but the set of global transformations could be isomorphic to a mere subset of the set of field configurations. The second reason has to do with the identification of the new term as a dynamical field, accompanied by an introduction of a term that governs its dynamics into the Lagrangian.

Both cases also share a notion of locality, which is not imposed but is rather an identified characterization of the evidence for the interaction-free theory. The applicability of inertial frames for the description of non-gravitational forces is perceived, in accordance with existing evidence, as a local matter, and therefore explained as such. The applicability of a preferred isospin convention is similarly explained as a local matter, determined by the local values of a conjectured bosonic field (that depends contingently by itself on its interaction with fermionic matter fields).

7 Foundational reflections

Having now motivated and presented the Methodological Equivalence Principle, we turn in this section to considering some of its foundational upshots. In §7.1, we discuss the construction of general relativity, the role of the Methodological Equivalence Principle in said construction, and the relation of that principle to the strong equivalence principle. In §7.2, we explain how our approach contributes to a 'dynamical' understanding of general relativity. In §7.3, we compare our approach with other means of constructing general relativity. In §7.4 we revisit the wider methodological and epistemological lesson from the theory.

7.1 From special to general relativity: the role of inertial frames

Einstein's [1905] appeal to the relativity principle in his development of the special theory of relativity tied the empirical content of the theory as well as its explanatory capacities to the notion of (what was for Einstein) an operationally-identified inertial coordinate system. This notion was born in the late 19th century attempts to replace Newtonian absolute space with an operationally-identified concept based on experience. Lange [2014], for example, defined in 1885 an inertial coordinate system using the (non-coplanar) trajectories of three free particles simultaneously projected from a single point. In this way, inertial structure is defined on the basis of observable properties.[22]

Having (supposedly) identified operationally the inertial frames of special relativity in this way, the construction of general relativity was then presented by Einstein as being based upon the principle (later to be known as the 'strong equivalence principle') stating that "[f]or infinitely small four-dimensional regions the theory of relativity in the restricted sense is appropriate, if the co-ordinates are suitably chosen" [Einstein, 1916, p. 154]. There are several problems with this approach. First, its motivation as a general principle may be unclear. Second, it is quite difficult to define the local validity of special relativity in a valid yet helpful way.[23] Third, even if we accept the requirement, it does not lead to general relativity (or to Riemannian geometry) in any straightforward way.[24]

The answer given here involves somewhat of a gestalt switch. Rather than seek to address the above issues regarding the strong equivalence principle from within the completed framework of general relativity (where we can agree the above issues constitute legitimate concerns[25]), we rather consider this principle from the point of view of the heuristics of theory construction. Then, the local (rather than global) validity of special relativity is first and foremost the appropriate conclusion from the evidence supporting special relativity, that boils down to localized experiments. Furthermore, the global *invalidity* of special theory is demonstrated in observations at large scale such as cosmological redshift. Further motivation comes from the identification of inertial frames with the freely falling ones, that follows from Einstein's equivalence principle. Note that our desideratum is weaker than most accounts of the local validity of special relativity: we would like to obtain a theory in which all special relativistic models are incorporated (approximately) as local models, but do not initially require that all of the local models of the new theory would be initially special relativistic.

This understanding of the local validity of special relativity as a statement about the empirical content of future theories, rather than as something to be derived from the formalism of those theories, gives rise to a methodological reconceptualization both of the principle of general covariance and of the equivalence principle, as means of achieving this local validity. Neither of these principle acts as a formal constraint in the construction. Instead, they are employed together in order to introduce the field that would render a particular set of coordinatizations as inertial within the set of all possible coordinatizations. This guarantees that in small regions of spacetime the equations of motion of matter will have their familiar special relativistic form to the extent that derivatives of the field $g_{\mu\nu}$ can be neglected. Thus, the introduction of the metric field using the Methodological Equivalence Principle amounts to introducing a minimal physical structure that accounts for this local validity.

With a bit of imagination, this suggested methodology can be read as an alternative history—as a possible path which Einstein himself could have followed in constructing and presenting his theory (arguably, a path not completely detached from some central ideas that served Einstein). In this case, one could speculate, the gauge argument might have been later suggested by direct analogy to Einstein's methodology, without the need for the mediation of Weyl's [1918] geometrical approach. Notably, this alternative history is antipodal to that presented by Stachel [2007], concerning the move from Newtonian gravitation theory to Newton-Cartan theory (the latter—like general relativity—being set on a manifold with curvature: see [Malament, 2012, ch. 4]). In brief, Stachel argues that *had* Newton known of the differential geometry of curved manifolds, he may have been able to arrive at the structure of Newton-Cartan theory. This strikes us as correct—however, it

[22] See [Barbour, 2001, ch. 12] for a review of attempts to identify operationally the notion of an inertial frame, and [Read, 2023a, ch. 1] for related discussion.

[23] For recent discussions on all these issues, see [Fletcher, 2020, Fletcher and Weatherall, 2023a,b, Linnemann et al., 2023, Read et al., 2018, Weatherall, 2020].

[24] See Torretti's [1983, p. 241] criticism of Reichenbach.

[25] See footnote 16, as well as related discussion below.

is clearly a very different point from that which we seek to make in this article, which aims to draw as much as possible from empirical and dynamical considerations. These can yield as an output a theory (in our case, general relativity) that can be read in terms of curved spacetime geometry. In this sense, we do not presuppose an enriched stock of geometrical resources *à la* Stachel.[26]

There may be an additional and deeper methodological lesson to be learned here. The methodology suggested here involves the assumption that the existence of inertial frames is an *explanandum*: the existence of these frames has to be accounted for in terms of underlying contingent physical structure. Inertial frames, it has been mentioned above, are sometimes (following *inter alios* Einstein) taken to have an operational definition. However, the construction of the theory does not appeal to this operational definition, but rather to a functional characterization of inertial frames as a theoretical concept, namely inertial frames are those in which the form of the laws simplifies maximally.

This characterization seems to echo inertial-frame functionalism, recently presented by Knox [2019] as a theoretically-informed approach in the philosophy of spacetime, and itself a variant of a more general approach of spacetime functionalism. Lam and Wüthrich [2021] have recently argued in the context of quantum gravity that spacetime functionalism has to be regarded in this project as an indispensable part of theory construction. The situation, they further argue, is different in general relativity, in which spacetime functionalism is merely one among several interpretational approaches. Our results show that in fact, the construction of general relativity can provide an exemplar for functionalist-spirited considerations in theory construction. The recipe that can be read from our construction consists of an identification of a theoretical concept (in our case the inertial frames), functional characterization of the concept based on its role in the theory (fixing the simple form of the laws; universality—see [Knox, 2013]), and then a construction of the minimalist physical, contingent realizer. In contrast to the conjectured role of functionalism in some quantum gravity theories, our recipe focuses on the functionalized concept, rather than on the realizer, and applies the mathematical considerations mainly in the scope of the theory that is already well-established (special relativity). This allows the new theoretical concept $g_{\mu\nu}$ to have unambiguous coordination with observation inherited from the operational characterization of the Minkowski metric.[27] Operational definitions of theorerical concepts therefore appear as a useful first step in theory construction, but incomplete without support by functional characterization.[28]

Finally, the notion of inertial frames that turns out to be useful in our construction is not only theory-laden, but also seem to depend on the choice of a particular formulation: coordinate representation of the dynamics of non-gravitational processes. However, this choice—one might worry [Fletcher, 2020, Fletcher and Weatherall, 2023b]—is problematically 'syntactic'; its meaning is lost in a coordinate-free formulation, and it has no imme-

[26] *Vis-à-vis* the move to Newton-Cartan theory: we claim that an alternative to Stachel's alternative history would be one in which Newton invoked the Methodological Equivalence Principle; we will, however, leave fleshing out the details here for another day. Cf. [Read and Teh, 2022].

[27] As in our account, Knox's notion of 'inertial frames' relates to coordinate systems in which the dynamical laws for material bodies take their simplest form: see [Knox, 2013]. For Knox, spacetime is that structure which identifies these inertial frames [Knox, 2019]. In general relativity, at least locally, the dynamical laws for material bodies take their simplest form in certain coordinate systems (related by Poincaré transformations); moreover, the metric takes its diagonal form in those coordinate systems. Thus, for Knox, in general relativity, the metric field qualifies as spatiotemporal. But Knox does not elaborate on why this coincidence arises: why is it that the metric field takes its diagonal form in the frames in which the laws simplify? This is what Read et al. [2018] call the 'second miracle of relativity'. However, if we are thinking about the metric as being constructed from dynamical considerations via the Methodological Equivalence Principle, then it seems that we have bridged this gap—of course the metric must take its form in the frames in which the dynamics simplify, because it was constructed to do just that!

[28] It is of course far from being obvious whether the role of inertial-frame functionalism in the theorizing can support spacetime functionalism as an ontological thesis on the nature of spacetime; this relate to issues such as multiple realizability and a 'hard problem' of spacetime [Le Bihan, 2021, Linnemann, 2021].

diate and clear definition in terms of structure on a manifold. What we learn, however, from the successful construction of gravitational coupling terms and kinematical structure using simple considerations that rely on the coordinate representation of inertial frames in special relativity, is that these entrenched representations *do* have intrinsic value; their being part of the syntax of special relativity does not mean that they are arbitrary, nor that they can discarded as merely 'syntactic'.[29]

7.2 Construction of a dynamical interpretation?

In our approach to arriving at the structure of general relativity, we have aimed to avoid presuppositions about the mathematical or geometrical structure of the constructed relativistic gravitational theory, and instead construct them using empirical extrapolations and dynamical considerations. Our motivation for that has been epistemological-methodological, rather than interpretational. However, the presented approach does demonstrate that Riemannian geometry is not heuristically indispensable in general relativity, and may therefore seem particularly appealing form a dynamical point of view. The aim of this subsection is to further explore the possible contribution of our construction to the dynamicist project and also to highlight ways in which it departs from Brown's dynamical account of general relativity [Brown, 2005].

Reflecting on the differences between geometrical and dynamical approaches to general relativity, Brown [2005, p. 150] identifies the two approaches with two possible chains of reasoning identified by Eddington [1930] that lead to the structure of the theory. The geometrical approach is identified with the familiar construction that sets off from the interval ds and the metric $g_{\mu\nu}$, constructs the Einstein tensor geometrically, and finally introduces the stress-energy tensor onto the other side of the equation. The dynamical approach takes a different—quite reversed—chain of reasoning, starting by describing matter using the stress-energy tensor, and then passing "to the interval regarded as the result of measurements made with this matter" [Eddington, 1930, 148].

The construction presented here demonstrates that the core of the logic of the first (perhaps more familiar) chain of reasoning can also be understood in dynamical terms. The new field $g_{\mu\nu}$ inherits its coordination with observations from the familiar field $\eta_{\mu\nu}$. It can similarly inherit the explanatory and interpretational baggage. Just as a geometrical reading of the Minkowski metric in the context of special relativity can naturally translate into a geometrical reading of the new field $g_{\mu\nu}$ (as a geometrical object with inherent chronogeometrical significance), a dynamical reading of the Minkowski metric (according to which the chronogeometric significance of the metric is the result of the special relativistic dynamical considerations) would translate a dynamical reading of $g_{\mu\nu}$.

The $g_{\mu\nu}$ field, which constitutes the basis for the new theory, can thus be understood as the metric field associated with spacetime geometry, but rather from dynamical considerations that apply to individual material fields via which we learn about the existence of this field. (Namely, it is the physical field that accounts for local inertial structure of the laws governing these fields.) The field equations are constructed later, based upon further dynamical considerations. Thus, if in the familiar picture the geometrical structure that underlies local inertial structure is replaced by spacetime geometry that is dynamically influenced by matter, here we see that this reasoning can work just as well and make perfect sense without using any geometrical vocabulary. The Methodological Equivalence Principle introduces $g_{\mu\nu}$ as the physical degrees of freedom that define local inertial frames. In the dynamical reading, it is the coupling to these degrees of freedom which leads to objects like rods and clock behaving in such a way as to be described in geometrical terminology.

[29]To be clear, this shouldn't be taken to imply that we don't take seriously the problems raised in [Fletcher, 2020, Fletcher and Weatherall, 2023b]—in fact, we agree that they are legitimate and serious concerns in the context of accounting for, say, the local validity of special relativity in the *completed* theory of general relativity.

Possible divergence from the dynamical view *à la* Brown [2005] concerns geodesic motion. Brown takes it that geodesic motion is "ultimately due to the way [...] $g_{\mu\nu}$ couples to matter" (p. 163). This is transparent in the construction presented in this article; our approach is dynamical in the sense that geodesic motion is accounted for rather than presupposed. The logic, however, is substantially different from that presented by Brown, according to whom the coupling of matter to the $g_{\mu\nu}$ field that underwrites geodesic motion is expressed by Einstein's field equations.[30] In the construction above, by contrast, geodesic motion and field coupling *precede* the field equations. The difference, of course, lies in that fact that we have been occupied with principles of theory *construction*, that are by definition non-deductive, whereas Brown (following Misner et al. [1973]) is concerned with analysing the structure of the *completed* theory, and with interpreting said theory. From the point of view of a proponent of a dynamical perspective on general relativity, one conceptual advantage to our approach is that it demonstrates that geodesic motion can make sense from a dynamical point of view even in cases in which the conditions of geodesic theorems which Brown has in mind (e.g. energy conditions: see [Malament, 2009]) fail to obtain.[31]

7.3 Relation to constructive approaches to general relativity

In this article, we have presented a way in which, beginning from special relativistic assumptions and the results of local experiments, one can arrive at the structure of general relativity. What we have discussed seems much in the spirit of 'constructive' (as opposed to deductive) approaches to general relativity.[32] Thinking about the theory in this way invites comparison with other 'constructivist' approaches to general relativity. Perhaps the most famous such approach is the 1972 'constructive axiomatization' of general relativity due to Ehlers et al. [2012] (henceforth 'EPS'). The idea of constructive axiomatics goes back to Reichenbach, who wrote that

> [i]t is possible to start with the observable facts and to end with the abstract conceptualization. [...] The empirical character of the axioms is immediately evident, and it is easy to see what consequences follow from their respective confirmations and disconfirmations. Such a *constructive* axiomatization is more in line with physics than a *deductive* one, because it serves to carry out the primary aim of physics, the description of the physical world. [Reichenbach, 1924/1969, p. 5]

In other words, the idea of Reichenbach's project of constructive axiomatisation is that one builds up the structure of the theory under consideration from axioms amenable to direct empirical test (rather than, as on a deductive approach, selected for, say, mathematical simplicity). In the EPS constructive axiomatisation, the differential, conformal, and projective structure of spacetime is built up from the trajectories of light rays and freely falling particles. In turn imposing an assumption that there is no second clock effect (which, recall, states that the rates of identical clocks can differ when transported to some end-point along two different paths) yields a Lorentizan spacetime structure—i.e., the kinematics of a spacetime theory using a familiar Lorentzian metric field $g_{\mu\nu}$ and associated Levi-Civita derivative operator.[33]

[30]The derivation is based on the following steps: The Einstein tensor $G_{\mu\nu}$ satisfies $G^{\mu}{}_{\nu;\mu} = 0$ as an identity. From the Einstein field equations it therefore must hold that $T^{\mu}{}_{\nu;\mu} = 0$. Matter moving along non-geodesics would violate this condition (or so the claim goes). For critical evaluation of this argument, bringing into the discussion famous 'geodesic theorems' such as that of Geroch and Jang [1975], see [Malament, 2009, Weatherall, 2011].

[31]In other words, what we mean to say here is this: the approach offered in this article can provide an explanation of geodesic motion in general relativity which obtains even when the conditions of certain geodesic theorems do not obtain. (Note that this is consistent with regarding the two different approaches as complimentary.)

[32]Note that what we mean by 'constructivist' here is not necessarily the same as what Einstein [1919] meant when he spoke of 'constructive theories'. For further discussion on this point, see [Adlam et al., 2022].

[33]For a detailed discussion of the EPS approach, see [Linnemann and Read, 2021, Adlam et al., 2022].

The EPS approach allows the intricate geometrical structures of general relativity to be defined operationally in a way that bears resemblance to Lange's operational definition of inertial frame mentioned in §7.1. This similarity highlights the difference from our construction: our starting point consists in not (merely) 'elementary' empirical facts, but rather in the empirical adequacy of the entire edifice of special relativity, its theoretical structures included. It thus assumes a detailed knowledge of local physics. In this sense, it is not fully constructivist, as it relies on antecedent theoretical inputs, rather than on elementary empirical facts. Although thereby rendering it less ambitious than the constructivist project, one could also argue that this makes our approach more realistic with respect to the nature of scientific knowledge, as it takes into account *all* knowledge of such local physics.

The difference between the two approaches goes further. In assuming antecedently that one has a grasp of the inertial motions, and so—by association—the inertial frames, EPS aim to use them as basis for more complicated structure. Our construction sets off from a similar observation, but the inertial frames then assume the status of an explanandum rather than an explanans, motivating thereby the introduction of the $g_{\mu\nu}$ field.

7.4 Lessons from general relativity

John Bell's [1976] paper on special relativity is commonly remembered either due to his spaceship paradox, or as a precursor of the dynamical approach. Bell's central message, however, is about theoretical continuity: we should prefer to construct and understand new physical theories using the concepts and content of old theories rather than completely new hypothetical conjectures.

The dramatic impact of the introduction of general relativity on the philosophy of science was, to a great extent, exactly about this issue of the epistemic status of hypothetical conjectures. Those seemed necessary in order to construct the abstract and intricate mathematical structures employed by the theory. One major possibility in both early and contemporary reflections is some revised version of apriorism. The major familiar empiricist alternative is motivated by verificationism, thus constructing mathematical concepts using operational definitions that work in the context of empirically-motivated axioms (§7.3). Later on, as appeal to abstract mathematics gradually came to dominate theoretical physics, epistemic and methodological reflections on theory construction became important to physics itself as well.

With the aim of emphasizing theoretical continuity with special relativity, the construction presented here is based on empirically-motivated methodological principles. Like the early Einstein, we began by using the equivalence principle in comparing inertial frames to frames that uniformly accelerate with respect to them. An apparently greater conceptual leap is from accelerating frames to all coordinate systems. This leap can be described in terms of applying the Methodological Equivalence Principle to the requirement for general covariance. We suggested that this apparent leap be understood not as an opaque mathematical constraint, but rather in close relation to experience, as an explanation of locally observed inertial phenomena in non-gravitational dynamics (namely, a way to introduce physical degrees of freedom required to define local inertial frames). This scheme transplants smoothly to applications of the gauge argument.

Reading this construction as a philosophical approach, it can be seen as an empiricist (though non-verificationist) alternative to constructive axiomatization, possibly highlighting a *tertium quid*; a golden mean between operationalism and apriorism. This approach suggests that understanding physical principles such as invariance principles and the principle of equivalence is an ongoing process of reflection on the development of physical theories and the interrelations between them. The understanding of such principles as methodological reflects the non-deductive way in which mathematical considerations play in theory construction. This reflective process emphasizes not only the continuity with earlier theories, but also with later ones such as extensions and modifications of general relativity (see end of §5) and gauge theories (§6). By aiming to bridge the content of different theories, this approach is more modest than approaches that aim to bridge between formal structure and empirical content, but arguably also closer to the actual nature of scientific development. A methodological understanding of familiar principles can, as we have seen, shed new light on philosophical puzzles related to them.

Compared with most existing philosophical views on general relativity, as well as with the methodological reflections of the later Einstein cited in the introduction, our approach motivates caution when appreciating the role of mathematical reasoning in physics. Mathematics is a vast land of unlimited possibilities. The case of general relativity may teach us that empirical considerations help us find our way in this landscape more that is usually appreciated.

Acknowledgements

We are very grateful to Yemima Ben-Menahem, Harvey Brown, and Noah Stemeroff for discussions on matters related to this paper and to two anonymous referees for their comments. J.R.'s work on this project was supported by a Leverhulme Trust Research Fellowship titled 'Measuring Spacetime'.

References

Emily Adlam, Niels Linnemann, and James Read. Constructive axiomatics in spacetime physics part II: Constructive axiomatics in context. 2022.

J. B. Barbour. *The Discovery of Dynamics: A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories*. Oxford University Press, 2001. ISBN 9780195132021. URL https://books.google.co.uk/books?id=pXc8DwAAQBAJ.

Julian B Barbour and Herbert Pfister. *Mach's principle: from Newton's bucket to quantum gravity*. Springer Science & Business Media, 1995.

John S. Bell. How to teach special relativity. *Progress in Scientific Culture*, 1, 1976. reprinted in Bell [1987].

John S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 1987.

Gordon Belot. Background-independence. *General Relativity and Gravitation*, 43(10): 2865–2884, 2011.

Clara Bradley. The non-equivalence of Einstein and Lorentz. *The British Journal for the Philosophy of Science*, 72(4):1039–1059, 2021. doi: 10.1093/bjps/axz014. URL https://doi.org/10.1093/bjps/axz014.

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

Harvey R. Brown and Dennis Lehmkuhl. Einstein, the reality of space and the action–reaction principle. In *Einstein, Tagore and the Nature of Reality*. 2016.

Harvey R. Brown and James Read. Clarifying possible misconceptions in the foundations of general relativity. *American Journal of Physics*, 84(5):327–334, 2016. doi: 10.1119/1.4943264. URL `https://doi.org/10.1119/1.4943264`.

Ernst Cassirer. Einstein's theory of relativity considered from the epistemological standpoint. *The Monist*, 32(3):412–448, 1922. doi: 10.5840/monist19223234.

S. Deser and R. P. Woodard. Nonlocal cosmology. *Phys. Rev. Lett.*, 99:111301, Sep 2007. doi: 10.1103/PhysRevLett.99.111301. URL `https://link.aps.org/doi/10.1103/PhysRevLett.99.111301`.

John F. Donoghue. General relativity as an effective field theory: The leading quantum corrections. *Phys. Rev. D*, 50:3874–3888, Sep 1994. doi: 10.1103/PhysRevD.50.3874. URL `https://link.aps.org/doi/10.1103/PhysRevD.50.3874`.

John F. Donoghue. Introduction to the effective field theory description of gravity, 1995. URL `https://arxiv.org/abs/gr-qc/9512024`.

John Earman. The implications of general covariance for the ontology and ideology of spacetime. In Dennis Dieks, editor, *The Ontology of Spacetime*, volume 1 of *Philosophy and Foundations of Physics*, pages 3–23. Elsevier, 2006a. doi: https://doi.org/10.1016/S1871-1774(06)01001-1. URL `https://www.sciencedirect.com/science/article/pii/S1871177406010011`.

John Earman. Two challenges to the requirement of substantive general covariance. *Synthese*, 148(2):443–468, 2006b. doi: 10.1007/s11229-004-6239-x.

Arthur Stanley Eddington. *The Mathematical Theory of Relativity*. Cambridge University Press, 1930.

Jürgen Ehlers, Felix AE Pirani, and Alfred Schild. Republication of: The geometry of free fall and light propagation. *General Relativity and Gravitation*, 44(6):1587–1609, 2012.

Albert Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der physik*, 4, 1905.

Albert Einstein. Über die vom Relativitätsprinzip geforderte Trägheit der Energie. *Annalen der Physik*, 328(7):371–384, 1907.

Albert Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen Der Physik*, 49(7):769–822, 1916. Translated as "The Foundations of General Relativity" in Lorentz et al., "The Principle of Relativity", Methuen and Company, Ltd. Reprinted by Dover Publications (1923).

Albert Einstein. Letter to Michele Besso. *Collected Papers of Albert Einstein*, 8:633, 1918.

Albert Einstein. What is the theory of relativity. *The London Times*, 1919. Reprinted in Einstein [1954b] pp.227-32.

Albert Einstein. Fundamental ideas and methods of the theory of relativity, presented in their development. In *The Collected Papers of Albert Einstein*, volume 7, chapter 31, pages 113–150. Princeton University Press, 1920. URL `https://einsteinpapers.press.princeton.edu/vol7-trans/129`. (English translation supplement).

Albert Einstein. Geometry and experience. In *The Collected Papers of Albert Einstein*, volume 7, chapter 52, pages 208–246. Princeton University Press, 1921. URL `https://einsteinpapers.press.princeton.edu/vol7-trans/224`. (English translation supplement).

Albert Einstein. Unified field theory of gravitation and electromagnetism. *Collected Papers of Albert Einstein*, 15:38–43, 1925.

Albert Einstein. Riemannian geometry retaining the concept of distant parallelism. *Collected Papers of Albert Einstein*, 16:210–214, 1928a.

Albert Einstein. New possibility for a unified field theory of gravitation and electricity. *Collected Papers of Albert Einstein*, 16:217–220, 1928b.

Albert Einstein. "on the present status of the general relativity theory. *The Yale University Library Gazette*, 6(1):7–10, 1931. Translated by Leigh Page.

Albert Einstein. On the method of theoretical physics. *Philosophy of Science*, 1(2):163–169, 1934. ISSN 00318248, 1539767X. URL `http://www.jstor.org/stable/184387`.

Albert Einstein. The problem os space, ether, and the field in physics. In *Ideas and Opinions*. Crown Publisher, New York, 1954a. Originally appeared in Mein Weltbild, Amsterdam: Querido Verlag, 1934.

Albert Einstein. *Ideas and opinions*. Crown Publisher, New York, 1954b.

Johannes Fankhauser and James Read. Gravitational redshift revisited: Inertia, geometry, and charge. 2023.

Samuel C. Fletcher. Apprxomate local Poincaré spacetime symmetry in general relativity. In Claus Beisbart, Tilman Sauer, and Christian Wüthrich, editors, *Thinking About Space and Time*. Birkhäuser, 2020.

Samuel C. Fletcher and James Owen Weatherall. The local validity of special relativity, part 1: Geometry. *Philosophy of Physics*, Nov 2023a. doi: 10.31389/pop.6.

Samuel C. Fletcher and James Owen Weatherall. The local validity of special relativity, part 2: Matter dynamics. *Philosophy of Physics*, Nov 2023b. doi: 10.31389/pop.7.

Michael Friedman. *Foundations of space-time theories: Relativistic physics and philosophy of science*, volume 113. Princeton university press, 1983.

Robert Geroch and Pong Soo Jang. Motion of a body in general relativity. *Journal of Mathematical Physics*, 16(1):65–67, 1975. doi: 10.1063/1.522416. URL `https://doi.org/10.1063/1.522416`.

Marco Giovanelli. Nothing but coincidences: the point-coincidence and einstein's struggle with the meaning of coordinates in physics. *European Journal of Philosophy*, 11, 2021.

Guy Hetzroni. Relativity and equivalence in Hilbert space: A principle theory approach to the Aharonov-Bohm effect. *Fundations of Physics*, 2020. `https://doi.org/10.1007/s10701-020-00322-y`.

Guy Hetzroni. Gauge and ghosts. *The British Journal for the Philosophy of Science*, 72(3): 773–796, 2021. doi: 10.1093/bjps/axz021.

Guy Hetzroni and Noah Stemeroff. Mathematical analogies in physics: The curious case of gauge symmetries. In Carl Posy and Yemima Ben-Menahem, editors, *Mathematical Knowledge, Objects and Applications: Essays in Memory of Mark Steiner*, pages 229–262. Springer International Publishing, Cham, 2023. ISBN 978-3-031-21655-8. doi: 10.1007/978-3-031-21655-8_11. URL `https://doi.org/10.1007/978-3-031-21655-8_11`.

Nick Huggett, Carl Hoefer, and James Read. Absolute and Relational Space and Motion: Post-Newtonian Theories. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition, 2022.

Michel Janssen and Jürgen Renn. *How Einstein found his field equations*. Birkhäuser, 2022.

Eleanor Knox. Effective spacetime geometry. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics*, 44(3):346–356, 2013. ISSN 13552198. doi: 10.1016/j.shpsb.2013.04.002.

Eleanor Knox. 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, 2019. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.2017.09.008. URL `https://www.sciencedirect.com/science/article/pii/S1355219817301247`.

Erich Kretschmann. Über den physikalischen Sinn der Relativitätspostulate. *Annalen Der Physik*, 53:575–614, 1917.

Vincent Lam and Christian Wüthrich. Spacetime functionalism from a realist perspective. *Synthese*, 199(Suppl 2):1–19, 2021. doi: 10.1007/s11229-020-02642-y.

Ludwig Lange. On the law of inertia. *The European Physical Journal H*, 39(2):251–262, 2014. Translation of "Über das Beharrungsgesetz" (1885), translated by Herbert Pfister.

Baptiste Le Bihan. Spacetime emergence in quantum gravity: functionalism and the hard problem. *Synthese*, 199(2):371–393, 2021. doi: 10.1007/s11229-019-02449-6. URL `https://doi.org/10.1007/s11229-019-02449-6`.

Dennis Lehmkuhl. Is spacetime a gravitational field? In Dennis Dieks, editor, *The Ontology of Spacetime II*, volume 4 of *Philosophy and Foundations of Physics*, pages 83–110. Elsevier, 2008. doi: https://doi.org/10.1016/S1871-1774(08)00005-3. URL `https://www.sciencedirect.com/science/article/pii/S1871177408000053`.

Dennis Lehmkuhl. Why Einstein did not believe that general relativity geometrizes gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46:316–326, 2014. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.2013.08.002. URL `https://www.sciencedirect.com/science/article/pii/S1355219813000695`.

Dennis Lehmkuhl. The equivalence principle(s). In *The Routledge companion to philosophy of physics*, pages 125–144. Routledge, 2021.

Niels Linnemann. On the empirical coherence and the spatiotemporal gap problem in quantum gravity: and why functionalism does not (have to) help. *Synthese*, 199(2):395–412, 2021. doi: 10.1007/s11229-020-02659-3. URL `https://doi.org/10.1007/s11229-020-02659-3`.

Niels Linnemann and James Read. Constructive axiomatics in spacetime physics part I: Walkthrough to the Ehlers-Pirani-Schild axiomatisation. 2021.

Niels Linnemann, James Read, and Nicholas Teh. The local validity of special relativity from an EFT-inspired perspective, 2023.

D. Lovelock. The Einstein tensor and its generalizations. *J. Math. Phys.*, 12:498–501, 1971. doi: 10.1063/1.1665613.

D. Lovelock. The four-dimensionality of space and the Einstein tensor. *J. Math. Phys.*, 13: 874–876, 1972. doi: 10.1063/1.1666069.

David B. Malament. On the status of the "geodesic law"' in general relativity. 2009.

David B. Malament. *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. Chicago: Chicago University Press, 2012.

Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973. ISBN 978-0-7167-0344-0, 978-0-691-17779-3.

Thomas Müller and Frank Grave. Catalogue of spacetimes. 2014. URL `http://go.visus.uni-stuttgart.de/CoS`.

John Norton. What was Einstein's principle of equivalence? *Studies in History and Philosophy of Science*, 16(3):203–246, 1985.

John D. Norton. General covariance and the foundations of general relativity: eight decades of dispute. *Reports on Progress in Physics*, 56(7):791–858, 1993. doi: 10.1088/0034-4885/56/7/001.

John D. Norton. Did Einstein stumble? The debate over general covariance. *Erkenntnis*, 42:223–245, 1995. ISSN 01650106. doi: 10.1007/BF01128809.

Lochlainn O'Raifeartaigh. *The dawning of gauge theory*. Princeton University Press, 1997.

J. Brian Pitts. Absolute objects and counterexamples: Jones–Geroch dust, Torretti constant curvature, tetrad-spinor, and scalar density. *Studies in History and Philosophy of Modern Physics*, 37:347–71, 2006.

Oliver Pooley. Substantive general covariance: Another decade of dispute. In Mauricio Suárez, Mauro Dorato, and Miklós Rédei, editors, *EPSA Philosophical Issues in the Sciences: Launch of the European Philosophy of Science Association*, pages 197–209. Springer Netherlands, Dordrecht, 2010.

Oliver Pooley. Background independence, diffeomorphism invariance, and the meaning of coordinates. In Dennis Lehmkuhl, editor, *Towards a Theory of Spacetime Theories*. Birkhäuser, 2015.

James Read. *Special Relativity*. Elements in the Philosophy of Physics. Cambridge University Press, Cambridge, 2023a. ISBN 9781009300612. doi: DOI: 10.1017/9781009300599. URL `https://www.cambridge.org/core/elements/special-relativity/97C5D3D0C577B3DF2609D571B18C811E`.

James Read. *Background independence in classical and quantum gravity*. Oxford University Press, 2023b.

James Read and Nicholas J. Teh. Newtonian equivalence principles. *Erkenntnis*, 2022. doi: 10.1007/s10670-021-00513-7. URL `https://doi.org/10.1007/s10670-021-00513-7`.

James Read, Harvey R. Brown, and Dennis Lehmkuhl. 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, nov 2018. ISSN 18792502. doi: 10.1016/j.shpsb.2018.03.001.

Hans Reichenbach. *The Theory of Relativity and a Priori Knowledge*. Berkeley: University of California Press, 1921/1965.

Hans Reichenbach. *Axiomatization of the Theory of Relativity*. Berkeley: University of California Press, 1924/1969.

Thomas Ryckman. *The Reign of Relativity: Philosophy in Physics 1915-1925*. Oxford University Press, 02 2005. ISBN 9780195177176. doi: 10.1093/0195177177.001.0001. URL https://doi.org/10.1093/0195177177.001.0001.

Simon Saunders. Rethinking newton's principia. *Philosophy of Science*, 80(1):22–48, 2013. doi: 10.1086/668881.

John Stachel. The story of Newstein or: Is gravity just another pretty force? In Michel Janssen, John D. Norton, Jürgen Renn, Tilman Sauer, and John Stachel, editors, *The Genesis of General Relativity*, pages 1962–2000. Springer Netherlands, Dordrecht, 2007. ISBN 978-1-4020-4000-9. doi: 10.1007/978-1-4020-4000-9_46. URL https://doi.org/10.1007/978-1-4020-4000-9_46.

J. L. Synge. *Relativity: the general theory*. North Holland Publishing Group, 1960.

Roberto Torretti. *Relativity and Geometry*. Dover Publications, 1983.

Jeroen van Dongen. *Einstein's Unification*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511781377.

David Wallace. Fields as bodies: a unified presentation of spacetime and internal gauge symmetry. 2015.

James Owen Weatherall. On the status of the geodesic principle in Newtonian and relativistic physics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 42(4):276–281, 2011. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.2011.09.004. URL https://www.sciencedirect.com/science/article/pii/S1355219811000566.

James Owen Weatherall. Fiber bundles, Yang–Mills theory, and general relativity. *Synthese*, 193(8):2389–2425, 2016. ISSN 15730964. doi: 10.1007/s11229-015-0849-3.

James Owen Weatherall. Two dogmas of dynamicism. *Synthese*, 199(S2):253–275, 2020. doi: 10.1007/s11229-020-02880-0.

Hermann Weyl. Gravitation and electricity. *Sitzungsber. Königl. Preuss. Akad. Wiss.*, 26: 465–480, 1918. English translation in O'Raifeartaigh [1997].

Hermann Weyl. Elektron und gravitation. i. *Zeitschrift für Physik A Hadrons and Nuclei*, 56(5):330–352, 1929a.

Hermann Weyl. Gravitation and the electron. *Proceedings of the National Academy of Sciences*, 15(4):323–334, 1929b.

Chen-Ning Yang. Einstein's impact on theoretical physics. *Physics Today*, 33(6):42–49, 1980.

Chen-Ning Yang and Robert L Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical review*, 96(1):191, 1954.