Einstein Studies
15

Claus Beisbart
Tilman Sauer
Christian Wüthrich
Editors

Thinking About Space and Time

100 Years of Applying and Interpreting General Relativity

Birkhäuser

Birkhäuser

Einstein Studies

Volume 15

Volume 1: Einstein and the History of General Relativity
Don Howard and John Stachel, editors

Volume 2: Conceptual Problems of Quantum Gravity
Abhay Ashtekar and John Stachel, editors

Volume 3: Studies in the History of General Relativity
Jean Eisenstaedt and A.J. Kox, editors

Volume 4: Recent Advances in General Relativity
Allen I. Janis and John R. Porter, editors

Volume 5: The Attraction of Gravitation: New Studies in the History of General Relativity
John Earman, Michel Janssen, and John D. Norton, editors

Volume 6: Mach's Principle: From Newton's Bucket to Quantum Gravity
Julian B. Barbour and Herbert Pfister, editors

Volume 7: The Expanding Worlds of General Relativity
Hubert Goenner, Jürgen Renn, Jim Ritter, and Tilman Sauer, editors

Volume 8: Einstein: The Formative Years, 1879–1909
Don Howard and John Stachel, editors

Volume 9: Einstein from 'B' to 'Z'
John Stachel

Volume 10: Einstein Studies in Russia
Yuri Balashov and Vladimir Vizgin, editors

Volume 11: The Universe of General Relativity
A.J. Kox and Jean Eisenstaedt, editors

Volume 12: Einstein and the Changing Worldviews of Physics
Christoph Lehner, Jürgen Renn, and Matthias Schemmel, editors

Volume 13: Towards a Theory of Spacetime Theories
Dennis Lehmkuhl, Gregor Schiemann, and Erhard Scholz, editors

Volume 14: Beyond Einstein
David E. Rowe, Tilman Sauer, and Scott A. Walter, editors

Volume 15: Thinking About Space and Time
Claus Beisbart, Tilman Sauer, and Christian Wüthrich, editors

More information about this series at http://www.springer.com/series/4890

Claus Beisbart • Tilman Sauer • Christian Wüthrich
Editors

Thinking About Space and Time

100 Years of Applying and Interpreting General Relativity

Birkhäuser

Editors
Claus Beisbart
Institute of Philosophy
University of Bern
Bern, Switzerland

Tilman Sauer
Institute of Mathematics
Johannes Gutenberg University of Mainz
Mainz, Germany

Christian Wüthrich
Department of Philosophy
University of Geneva
Genève, Switzerland

ISSN 2381-5833 ISSN 2381-5841 (electronic)
Einstein Studies
ISBN 978-3-030-47781-3 ISBN 978-3-030-47782-0 (eBook)
https://doi.org/10.1007/978-3-030-47782-0

Mathematics Subject Classification: 83-03, 01-06, 03A05

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Einstein Studies Series Preface

Einstein Studies was launched in 1989 under the joint editorship of Don Howard and John Stachel, the founding editor of *The Collected Papers of Albert Einstein* and the Director of Boston University's Center for Einstein Studies, which served as the administrative home for the *Einstein Studies* series. The series was envisioned as a companion to the Einstein Papers Project, then also housed at Boston University, as a venue for the publication of scholarship relating to all aspects of the life and work of Albert Einstein, and as a tool for engendering and supporting an expanding community of scholars, especially younger scholars, working on such topics. *Einstein Studies* also aimed to be broadly interdisciplinary, featuring not only work on the history of science and technical work in physics, but also philosophy of science and social science perspectives on physics and its cultural embedding.

At the time of John Stachel's resignation as co-editor in 2017, the *Einstein Studies* series had published a total of thirteen volumes on topics ranging from *Einstein and the History of General Relativity* and *Einstein's Formative Years* to *Mach's Principle* and *Einstein Studies in Russia*. Included among those thirteen volumes is the rich and important volume one of Stachel's own collected papers, *Einstein from 'B' to 'Z.'* It would not be immodest to say that the series has realized its early ambition by helping to build what is now a thriving international scholarship focused on Einstein, a body of scholarship that is exemplary in its technical sophistication, historical depth, philosophical acuity, and cultural contextualization.

With Diana Kormos Buchwald's assumption of the role of co-editor, the vital connection between the *Einstein Studies* series and the Einstein Papers Project, which she directs, is reaffirmed. With the addition of a distinguished, international, editorial advisory board, the series is poised to play an even more prominent role in fostering the further expansion and enhancement of Einstein scholarship, with, again, special attention to nurturing each new generation of younger scholars as they enter the field. We eagerly invite proposals covering every part of the subject terrain.

Notre Dame, IN — Don Howard
Pasadena, CA — Diana L. Kormos-Buchwald

Introduction

The year 1915 saw one of the most momentous events in the history of science: on 25 November of that year, Einstein completed his decade-long search for a general theory of relativity by presenting his final gravitational field equations to the Prussian Academy for publication in its Proceedings. But despite Einstein's alleged "heart palpitations" that he said he experienced when he succeeded to compute the correct value of Mercury's perihelion, the quest for a new theory of gravitation was hardly answered yet. No solutions to the field equations were known, and the implications of the theory both for physics proper and for our philosophical understanding of physics were largely in the dark.

The publication of Einstein's field equations immediately gave rise to pioneering research. Only a few weeks after publication of Einstein's field equations, Karl Schwarzschild found the first nontrivial solution to these equations. Einstein himself continued to ponder the question of whether the theory faithfully reflected his underlying heuristic ideas, in particular, concerning what he began to call "Mach's Principle." During an exchange with the Dutch astronomer Willem de Sitter, he developed the first relativistic world model and thus laid the foundations for modern theoretical cosmology. And in 1919, Sir Arthur Eddington confirmed Einstein's prediction for the reflection of light during an eclipse, a feat that immediately proved decisive for Einstein's popularity in the broader public.

Since the theory had deep and far-reaching implications for our understanding of space and time, the discovery of the field equations was also followed by intense philosophical debate. Prominent proponents of positivism and neo-Kantianism, e.g., Moritz Schlick, Rudolf Carnap, and Ernst Cassirer, made contributions of lasting importance.

In view of these developments, it seems fair to say that, in November 1915, the theory had only just arrived but was not fully developed, yet; what we know and value as the theory of general relativity really only came into being in the aftermath of the discovery of the field equations, triggered by this event and building on it.

The aim of this volume therefore is to reconsider Einstein's theory from the perspective of its immediate and later reception in physics and philosophy. We aim at an integrated understanding of Einstein's theory by combining perspectives from both history and philosophy.

The contributions to this volume were originally presented at a conference held in Berne, from 12 to 14 September 2017, funded by the UBS Culture Foundation, the Tomalla Foundation, the Swiss National Science Foundation, the Albert Einstein Center for Fundamental Physics, and the Institute of Philosophy at the University of Bern. We would like to express our gratitude for the generous support of these institutions. We also thank the anonymous referees for this collection for their voluntary service, as well as the Springer team and Chris Eder for their continued support and the smooth cooperation.

Since many contributions to this volume integrate historical and philosophical perspectives (which we take to be very fruitful), they do not divide themselves naturally into categories that stress the viewpoint of one of either discipline, respectively. Nevertheless, some contributions are more historically oriented, while others focus more on philosophical aspects. We have thus organized the volumes by starting with the more historical contributions and then moving forward towards more philosophically oriented chapters.

The first chapter of the volume puts our very own conference into a historical perspective. **Claus Kiefer** takes a look back at another conference held in 1955 in Bern, which in some ways marks the beginning of the renaissance of the theory. Taking place only weeks after Einstein's death, the 1955 Bern conference can be considered the true beginning of a regular international conference series "General Relativity and Gravitation," which is still ongoing and which led to the institutionalization of an ever-growing and active research community in general relativity. Kiefer, himself a physicist working in general relativity and quantum gravity, offers an assessment of the scientific content of the 1955 Bern conference. He comments on the papers given then and puts them into a larger perspective of later developments, singling out developments in classical general relativity, cosmology, and the very early work on quantum gravity.

John D. Norton goes further back in time and traces Einstein's path from the special theory of relativity to the discovery of the general theory. The story of Einstein's discovery of the gravitational field equations has been subject of intensive scrutiny and the story has been told many times, also in pioneering work by Norton himself. Nevertheless, neither has consensus been reached about all aspects of this nor have some notorious dark points been fully clarified yet. Norton addresses one such point which actually pertains to the heart of Einstein's intellectual journey. His point of departure is the unclear role that the principle of equivalence played in Einstein's thinking at the time, given the fact that the precursor theory of general relativity did not play out well on this explicit heuristics. Norton goes back to Einstein's early steps towards a static theory of gravitation and shows how a second tier of heuristics, viz. energy–momentum conservation, emerges as a powerful constraint that for some time overruled the more explicit demands of general equivalence.

One of the most important early applications of general relativity pertained to cosmology. Einstein's 1917 landmark paper "Cosmological Considerations in the

General Theory of Relativity" ("Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie") established the first relativistic model of the Universe. In his chapter, **Cormac O'Raifeartaigh** examines Einstein's model from both a historical and a philosophical perspective. He stresses the important role that Mach's principle played for the introduction of the model. Einstein had relied on the principle when introducing his very theory and used it to rule out certain boundary conditions since he took them to be incompatible with the principle. Einstein finally came up with a spatially closed model of the Universe, the static character of which could only be maintained by introducing a cosmological constant in the field equations. In this respect, O'Raifeartaigh notes a slight mathematical inaccuracy that may have influenced Einstein's understanding of the new term. As O'Raifeartaigh further stresses, in his 1917 paper Einstein did not test his models against observations, although he could have done so; nor did he consider the stability of his world model.

The instability of his world model later became one of the prime reasons for Einstein to give it up. As **C. D. McCoy** points out, considerations of stability have also played an important role in inflationary cosmology. One reason to believe that the early Universe was inflated by more than twenty orders of magnitude is that the Universe would otherwise be unstable with respect to its flatness: Small deviations from flatness would increase and lead to a Universe that is curved at large scales. It thus seems that stability is a key standard on cosmological models. In his chapter, McCoy spells doubts on this idea. After a close analysis of the instability of the Einstein world and of a flat cosmological model, he discusses stability from the perspective of dynamical systems theory. He argues that there are pragmatic reasons for assuming stability, but that an outright rejection of any unstable model is not justified.

A decisive application of any theory of gravity is the two-body problem; in general relativity, it plays a particularly important role because of its connection to the problem of motion. **Galina Weinstein**'s chapter turns to Einstein's work on the two-body problem, which resulted in a particular solution to the gravitational field equations, the so-called Einstein–Rosen bridge, published in 1935. At the same time, he did also research on the thought experiment of the famous "EPR"-paper by Einstein, Rosen, and Podolsky (1935). So might the Einstein–Rosen bridge have served as a heuristic guide for Einstein in developing the quantum mechanical thought experiment? Weinstein lays out the historical evidence for this conjecture and contends that although we have no conclusive evidence for the claim that the EPR paradox would have been inspired by the Einstein–Rosen bridge, it is not an unlikely scenario.

With the formulation of non-Euclidean geometry in the 19th century, and with renewed urgency once general relativity came to be accepted, the "problem of space," i.e., the challenge of determining the geometrical structure of physical space, arose. **Neil Dewar** and **Joshua Eisenthal** present Weyl's solution to the problem of space, which he presented in 1922 in a series of lectures in Barcelona. Weyl started from the idea that the demand for free motion by a rigid body to be "localized" implied that acceptable geometries would just be those described by a metric whose form is infinitesimally Pythagorean, i.e., can be written as some positive-definite quadratic form. These “Riemannian” metrics have unique symmetric affine connections compatible with them and so are associated with a unique notion of parallel transport. Dewar and Eisenthal then show how Weyl’s insight and the concepts he introduced can be used to illuminate the contemporary debate on whether spacetime is represented in general relativity just by the bare topological manifold (while the metric counts as additional physical field) or by the manifold together with the metric field defined on it. Weyl offers a via media between these poles: the “nature” of the metric is part of spacetime’s intrinsic, fixed essence, while its “orientation” is only a posteriori given and contingently determined by the material content of the world and consequently not part of spacetime. Dewar and Eisenthal feature Weyl’s view as a contribution that is still significant to the question of what it is to be physical space or spacetime.

In the following chapter, **Ryan Samaroo** addresses the status of the principles underlying general relativity, among them the principle of equivalence. According to Michael Friedman, who has adopted a broadly Kantian perspective, physical theories have several layers. Some principles of a theory are constitutive because they open the conceptual space for certain possibilities and thus define a framework for empirical investigation. Other principles presuppose this framework and shrink the possibilities on empirical grounds. This outlook echoes the Kantian distinction between a priori and a posteriori and contrasts with a Quinean picture in which the principles of a theory do not admit of a principled distinction in, e.g., a priori and a posteriori. Friedman has prominently illustrated his approach to physical theories using Einstein’s general relativity. For Friedman, the principle of equivalence is a constitutive one, while the field equations are of the other type. Samaroo’s main aim in his contribution is to defend a broadly Friedmanian account of general relativity against objections. One objection, for instance, has it that there is no unique way to make the distinction between both kinds of principles. Samaroo argues that some of the objections are based upon misunderstandings. But he concedes to the critics that the principle of equivalence should not be regarded as constitutive.

In the next contribution, **Niels Linnemann** addresses the question to what extent the theory of general relativity is in need of interpretation. That question famously arises for the theory of quantum mechanics where the measurement problem makes a straightforward interpretation of the theory impossible. Taking as his starting point work by Erik Curiel who distinguished between three kinds of interpretation, concrete, categorical, and meta-linguistic, Linnemann proposes a further refinement of this analysis, suggesting a new category of interpretation which he calls “qualificatory interpretation.” His claim is that the general theory is, in fact, in need of this kind of interpretation, and his two key arguments pertain to the question of chronometric interpretation and the problem of what it means to assign thermodynamic notions like entropy to solutions of the gravitational field equations like the surface area of a black hole. In his chapter, Linnemann also explores how a concrete or a categorical interpretation of GR can guide the search after a successor theory.

James Read’s contribution discusses the so-called dynamical view of spacetime theories, which is contrasted with the more standard “geometric” view. In his characterization, the geometric view takes the metric field to be an autonomous physical entity, which constrains the dynamics of matter fields. The geometric view is thus a brand of substantivalism. In contrast, Read defines the dynamical view as relegating the metric field from the status of an ontologically autonomous entity to a mere codification of the symmetry properties of the dynamics of present matter fields, and so a form of relationalism. This holds at least at the level of special relativity; once we move to general relativity, advocates of the dynamical view must concede that the metric is an independent entity, which cannot be reduced to matter fields and their dynamics because of the gravitational degrees of freedom postulated in general relativity. This is ultimately the reason why defensible versions of the dynamical view collapse into a form of the geometric view, at least of the "qualified" kind, at the level of general relativity. Using Jim Weatherall's recent writings on the geodesic principle as a foil, Read closes in arguing that viable versions of the dynamical and the geometric views on general relativity not only converge but also naturally lead to a form of spacetime functionalism.

In analogy to Newton's first law, the geodesic principle of general relativity identifies the possible trajectories of free, i.e., "inertially" moving, massive point particles with time-like geodesics. Following a long tradition of debating the status of the geodesic principle in general relativity as either an independent postulate or a theorem following from basic assumptions, **James Owen Weatherall** investigates the possibility of finding a formulation of the geodesic principle that does not make problematic reference to point particles and analyses the extent to which such a reformulated principle could obtain the status as theorem in general relativity. Weatherall first discusses two ways in which one might prove a geodesic principle, none of which is satisfactory from a physical point of view. He then proposes a new approach based on "tracking," which combines the two approaches and permits overcoming, to a significant degree, the shortcomings of both. Tracking uses constructions allowing us to capture the idea that matter fields as encoded in the stress-energy tensor T_{ab} are as concentrated as one likes for a finite duration as long as one likes near a given curve. It also permits a reformulation of the geodesic principle according to which the energy–momentum tensors associated with solutions to source-free matter field equations only track time-like or null geodesics. As intended, the reformulated principle, which can be established (almost, i.e., modulo the satisfaction of the dominant energy condition) as a theorem in general relativity, associates certain "inertial" motions of physical bodies with a geometrically elite class of curves—the geodesics.

As mentioned before, one of the principles that led Einstein to his theory was Mach's principle. According to the principle, roughly, inertial motion is not determined by the nature of some absolute space, but rather by the distribution of matter. But what exactly does it mean to say that the distribution of matter determines how inertial motion is like? This is the question that **Antonio Vassallo** and **Carl Hoefer** address in their chapter. Their focus is on the so-called rotational frame-dragging effects that manifest the Machian nature of general relativity. One example is the Einstein–Lense–Thirring effect in which the rotation of a gyroscope is determined by the surrounding mass distribution. As Vassallo and Hoefer point out, it is not attractive to conceive of this determination in terms of causation. The reason is that a change in the mass distribution affects inertial motion in an immediate way; while causal influence is most often thought to propagate with the speed of light at most. To develop an alternative metaphysical picture of the dependence relation, Vassallo and Hoefer use structural equation modeling and develop a precise model of the way in which the matter distribution impacts on the inertial motion. The question of how the determination relation is best understood then turns on the understanding of the structural equation framework. One answer that they suggest is that the relation is halfway between causation and grounding.

How does general relativity explain the success of special relativity in dealing with physical phenomena not involving gravity? One would surely expect general relativity—a more encompassing theory than special relativity—to subsume special relativity and explain both its success and its shortcomings. In his contribution, **Samuel C. Fletcher** uses the notion of an "approximate local spacetime symmetry" he developed elsewhere in order to offer central aspects of such an explanation. Applying this notion, it can be shown that every general-relativistic spacetime is approximately locally Poincaré symmetric. Since Minkowski spacetime is (precisely and globally) Poincaré symmetric, the approximate symmetry of relativistic spacetimes connects general relativity to special relativity. Fletcher notes that, although a keystone, this fact is insufficient by itself as it does not in general account for why the observable behavior of matter fields in generic spacetimes is well approximated by those of corresponding fields in Minkowski spacetime.

Institute of Philosophy, University of Bern, Bern, Switzerland — Claus Beisbart
Institute of Mathematics, Johannes Gutenberg University of Mainz, Mainz, Germany — Tilman Sauer
Department of Philosophy, University of Geneva, Genève, Switzerland — Christian Wüthrich

Contents

List of Contributors

Neil Dewar Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, München, Germany

Joshua Eisenthal Department of Philosophy, University of Pittsburgh, Pittsburgh, PA, USA

Samuel C. Fletcher University of Minnesota, Minneapolis, MN, USA

Carl Hoefer Universitat de Barcelona, Gran Via de les Corts Catalanes, Barcelona, Spain
ICREA, Barcelona, Spain

Claus Kiefer Institute for Theoretical Physics, University of Cologne, Köln, Germany

Niels Linnemann University of Bremen, Bremen, Germany

C. D. McCoy Underwood International College, Yonsei University, Seoul, South Korea

John D. Norton Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

Cormac O'Raifeartaigh School of Science and Computing, Waterford Institute of Technology, Waterford, Ireland

James Read Pembroke College, Oxford, UK

Ryan Samaroo Somerville College, Oxford, UK

Antonio Vassallo International Center for Formal Ontology, Warsaw University of Technology, Warsaw, Poland

James Owen Weatherall Department of Logic and Philosophy of Science, University of California, Irvine, CA, USA

Galina Weinstein Department of Philosophy, University of Haifa, Haifa, Israel
Israel and Interdisciplinary Center (IDC), Herzliya, Israel

Space and Time 62 Years After the Berne Conference

Claus Kiefer

Abstract In 1955, an international conference took place in Berne in which the state of relativity theory and its possible generalizations were presented and critically discussed. I review the most important contributions to that conference and put them into the perspective of today's knowledge about the nature of space and time.

1 Historical Context

From July 11 to July 16, 1955, a conference took place in Berne celebrating the fiftieth anniversary of the special theory of relativity. This was perhaps the first international conference devoted to an overview of relativity theory, its ramifications, and applications. The main goal of that conference was neither historical nor was it restricted to special relativity; in fact, most of the topics deal with general relativity and its generalizations, both in classical and quantum directions, along with cosmology and with mathematical structures. The list of participants contains an impressive number of famous figures together with a selection of young scientists.[1] The Proceedings of that conference were published in 1956 and contain most of the presented talks together with a record of the discussions (Mercier and Kervaire, 1956). In my contribution, I will heavily rely on these Proceedings, the title page of which is displayed in Fig. 1.

The Berne Conference was later known as the GR0 Conference, where the numbering refers to the series of conferences organized by The International Society on General Relativity and Gravitation (GRG).[2] This society was founded

[1] One of the younger participants, Walter Gilbert, was awarded the 1980 Nobel Prize in chemistry.
[2] See http://isgrg.org.

C. Kiefer (✉)
Institute for Theoretical Physics, University of Cologne, Köln, Germany
e-mail: kiefer@thp.uni-koeln.de

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_1

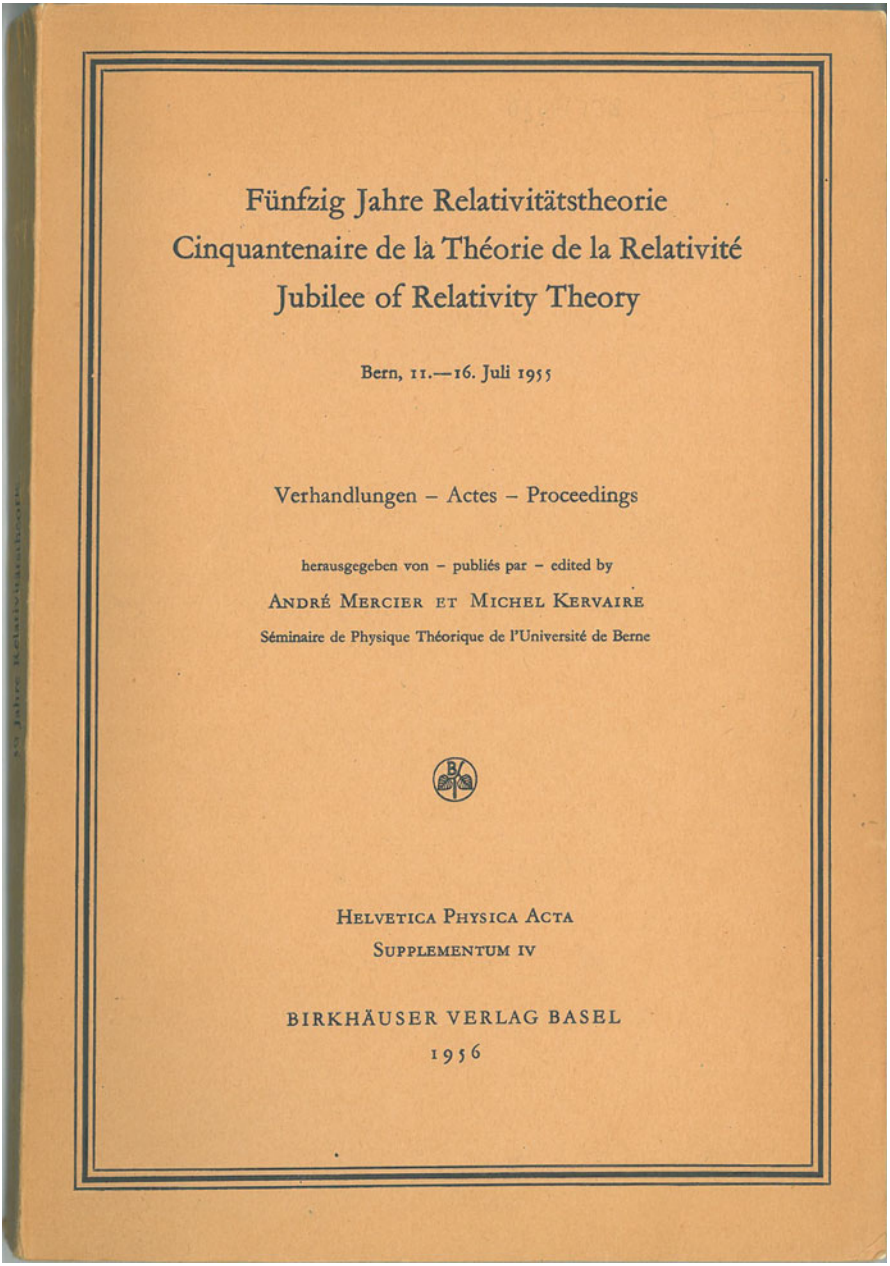


Fig. 1 Title page from the Proceedings of the Bern Conference 1955. © Birkhäuser Verlag, Basel. Photo by the author

Fig. 2 Contemporary look of the building at Sidlerstrasse 5 (photo: M. A. Vogt (cutout); from Burgerbibliothek Bern, FPa.9, Bl. 43, http://katalog.burgerbib.ch/detail.aspx?ID=105643)

at the GR6 conference, which took place in Copenhagen in 1971, and grew out of the International Committee on General Relativity and Gravitation, which was responsible for the earlier conferences. Several of the GRG Presidents were, in fact, participants of the Berne Conference, among them were Christian Møller, Nathan Rosen, Peter Bergmann, and Yvonne Choquet-Bruhat.

The conference was organized by the Seminar for Theoretical Physics of the University of Bern, located at the Institute for Exact Sciences at Sidlerstrasse 5. The President of the Organization Committee was Wolfgang Pauli (Zürich), the Scientific Secretary André Mercier from Berne. The lectures themselves took place in the lecture hall of the Natural History Museum. Figure 2 shows the building at Sidlerstrasse at the time of the conference. It was there that Einstein delivered his lectures as a *privatdozent* in Berne. These were the lectures on molecular theory of heat (*Molekulare Theorie der Wärme*) in the summer term 1908 (with three attendants, which were his friends), and on the theory of radiation in the winter term 1908/09 (with four attendants, even including a student), see Fölsing (1994, p. 274).

Between 1959 and 1963, the new building shown in Fig. 3 was erected, designed by the architect couple Hans and Gret Reinhard. It clearly demonstrates the new spirit of the day.[3]

In his foreword to the Proceedings, the secretary André Mercier made the following interesting remarks on the theory of relativity:

> The theory of relativity marks all in all one term: it is the achievement of a physics of cartesian spirit that gives an account of the phenomena by figures and by motions . . . One

[3] See http://unibe.ch/university/portrait/history/.

Fig. 3 Present look of the building at Sidlerstrasse 5 (photo and ©: Chr. Schläppi)

> hears today the saying that we live in the atomic era. Should we not also speak of the relativistic era?[4]

This, of course, alludes to the fact that the theory of relativity was not very popular in the 1950s, apparently overshadowed by quantum theory and its applications to atomic and nuclear physics.

Of interest is also the welcome speech of a local politician, a certain Dr. V. Moine, *Directeur de l'instruction publique du Canton de Berne*. He reflects the philosophical spirit of the conference's location:

> This city, enclosed like a jewel by the crown of the river Aare, with its military and political past which made it the head of the old Switzerland, practical and empirical as a farmer's wife, has always valued the positive and immediate higher than the theoretical. Its symmetric streets, its order, the equilibrium which is brought out by its buildings, the traditional caution of its laws make it more a city of Aristotle than of Platon. . . .[5]

(The building shown in Fig. 3 perhaps also gives testimony of this Aristotelian attitude.) This practical spirit is also seen in the organization of the conference. Three languages (English, French, and German) are used interchangeably, and in the discussions they are often mixed in an interesting way; an example is the discussion

[4]This is my translation from the original French which reads: "La Théorie de la Relativité marque en somme un terme: elle est l'achèvement d'une physique d'esprit cartésien rendant compte des phénomènes par figures et par mouvements. . . . On entend dire aujourd'hui que nous vivons à l'ère atomique. Ne pourrait-on aussi bien parler de l'ère relativiste?" (Mercier and Kervaire 1956, p. 19).

[5]This is my translation from the original French which reads: "Cette ville, enserrée comme un joyau dans la couronne de l'Aar, au passé militaire et politique qui fit d'elle la tête de la vieille Suisse, pratique et empirique comme une paysanne, a toujours plus apprécié le positif et l'immédiat que le théorique. Ses rues symétriques, son ordonnance, l'equilibre qui se dégage de ses édifices, la prudence traditionelle de ses lois en font plus une cité d'ARISTOTE que de PLATON. . . . " (Mercier and Kervaire 1956, p. 25).

after Bergmann's talk with its mélange of the three languages (Mercier and Kervaire 1956, pp. 95/96).

Albert Einstein had been invited to this conference, but he died on April 18, 1955, three months before the Berne Conference. Anyway, he had not envisaged to attend the meeting. In his reply to a letter of invitation by Louis Kollros,[6] Einstein had written:

> We two are no spring chickens anymore! As for me, I cannot think about a participation. ...[7]

It is left entirely to our imagination to figure out what would have happened if Einstein had been able to attend the Berne meeting.

2 Classical General Relativity and Beyond

The year 1955 marked the 40th anniversary of Einstein's general theory of relativity. Since it was difficult in those years to test the theory empirically beyond the classic tests (redshift, light deflection, perihelion motion), much attention was focused on theoretical and mathematical developments. This concerned, in particular, the structure of the Einstein field equations, notably the initial value problem and the problem of motion. As for the former, two of the main figures, André Lichnerowicz and Yvonne Choquet-Bruhat (at that time Fourès-Bruhat) were present at the meeting. As for the latter, Leopold Infeld was the main figure who was present.

The well-posedness of the initial value problem (Cauchy problem) is of great importance. Only if there exist initial data, that is, data on a three-dimensional hypersurface that determine the evolution according to the Einstein equations uniquely, can one use the theory to predict physical processes, for example, the emission of gravitational waves from coalescing compact objects. Today, well-posedness is generally granted as established, see, for example, Isenberg (2014). It is a key ingredient in numerical relativity.

By the time of the Berne Conference, a first theorem on the initial value problem had already been proven by Choquet-Bruhat in 1952. A more general theorem was proven in a 1969 paper of Choquet-Bruhat and Robert Geroch, see Isenberg (2014) and Chruściel and Friedrich (2004) for a detailed discussion and references.[8] The theorem proven by Choquet-Bruhat and Geroch can be stated as follows (see Isenberg 2014, p. 307):

[6] Louis Kollros (1878–1959) was a Swiss mathematician; from 1909 to 1948 he was a professor at ETH Zürich.

[7] This is my translation from the original German which reads: "Wir sind beide keine Jünglinge mehr! Was mich betrifft, so kann ich nicht an eine Beteiligung denken. ... " (Mercier and Kervaire 1956, p. 271).

[8] The subtitle of the volume Chruściel and Friedrich (2004) in fact reads "50 Years of the Cauchy Problem in General Relativity".

Theorem *For any smooth set of initial data* (h_{ab}, K_{cd})*, where* h_{ab} *is the three-metric and* K_{cd} *is the extrinsic curvature (second fundamental form), on a specified three-manifold which satisfies the vacuum constraint equations, there exists a unique (up to diffeomorphism) maximal globally hyperbolic development.*

Further developments are discussed in the above cited references. One concerns the extension to the non-vacuum case: the theorem also holds, for example, for the physically relevant case of the Einstein–Maxwell theory. On the mathematical side, it has been shown that the required degree of regularity can be weakened. Other developments concern the stability of Minkowski spacetime under long-time evolution, the stability of de Sitter space, and investigations on the cosmic censorship conjecture. The latter conjecture—in its weak form stating that the singularities arising from gravitational collapse cannot influence future null infinity—was formulated by Roger Penrose in 1969. Most of these later investigations made heavy use of the global methods (Penrose–Carter diagrammes) developed in the 1960s, which were unavailable at the time of the Bern Conference.

Currently, there is much interest in classical generalizations of general relativity; concrete examples are the $f(R)$ theories, where R is the Ricci scalar. Whether those theories also enjoy a well-posed initial value problem is far from clear. It is thus too early to make statements about the range of validity for those theories. It is imaginable that they can only be applied in more restricted situations and not, for example, to the non-perturbative treatment of gravitational wave emission.

The subject of gravitational waves, which after their first direct detection in 2015 is of central importance today, received little attention in 1955. Nathan Rosen, in his talk, basically reviewed his work with Einstein of 1937 in which they had expressed doubts about the existence of gravitational waves in the full non-linear theory. According to them, the plane wave solutions of the linearized theory do not correspond to any exact solution of the full theory. Today we know, of course, that they were in error and that gravitational waves indeed exist.

The experimental situation with general relativity was not in a good shape at the time of the conference. Still, the state of the art of the two classic tests concerning gravitational redshift and light deflection (as well as the state of cosmology, see below) was addressed. The gravitational redshift (time dilation) is, for a constant field with gravitational acceleration g, given by the standard formula

$$\frac{\Delta\nu}{\nu} = \frac{gh}{c^2}. \tag{1}$$

In his contribution, Robert Trumpler reported on recent (from 1954) observations in the spectral lines of the white dwarf 40 Eridani B and gave a list with the redshifts measured for 18 other stars. The historic experiments by Pound and Rebka, determining the redshift in a laboratory experiment using the newly discovered Mössbauer effect, were still 4 years ahead. Those new types of experiments are of much higher accuracy than the stellar observations which have thus lost their significance. Today, the gravitational redshift effect is part of everyday life, for example through the use of the Global Positioning System (GPS). For a detailed discussion of the current experimental situation, see Will (2014); frequency shifts have been measured over a height of 1/3 of a metre.

The second classic test discussed at Berne was light deflection. For a grazing ray near the Sun, the deflection angle is given by

$$\delta = \frac{4GM}{Rc^2} \equiv \frac{2R_S}{R}, \tag{2}$$

where R is the solar radius, M the solar mass, and R_S is the Schwarzschild radius. It is convenient to parametrize this effect by a post-Newtonian parameter γ, which assumes the value $\gamma = 1$ in general relativity,

$$\delta \approx \frac{1+\gamma}{2} 1.''7505. \tag{3}$$

The first observations were the famous ones performed at Sobral and in Principe on the occasion of the solar eclipse on May 29, 1919. The accuracy there was about 30%. In his talk at Berne, Trumpler reported about results from other eclipses, those of 1922, 1929, 1947, and 1952. The accuracies there were not much better than in 1919. Today, light deflection has been confirmed to an accuracy of 0.01% (Will, 2014). This is mainly due to the development of very-long-baseline radio interferometry (VLBI). Still, observations during the total solar eclipse of August 21, 2017 in the USA have led to a value for the light deflection of 1.7512 arcsec, with an uncertainty of only 3% (Bruns, 2018).

As is evident from Will (2014), the progress in experimentally testing general relativity since 1955 has been tremendous, and the status of the theory in this regard is similar to elementary particle physics.

There have been many developments since that no one could have imagined in 1955. These concern, in particular, the field of relativistic astrophysics, which more or less started in 1963 with the discovery of the first quasar 3C 273 by Maarten Schmidt. For the study of active galaxies, neutron stars, and black holes, general relativity has proven indispensable. The very concept of a black hole was not understood in 1955 and played no role at the conference. Today, we can gain insight into the coalescence of black holes and neutron stars by investigating the gravitational waves they emit. A single black hole can be studied by its influence on the surroundings; a prominent example is the supermassive black hole in the centre of our Milky Way (Eckart et al., 2017).

At the conference, some interest was also devoted to classical generalizations of Einstein's theory. In the last few decades of his life, Einstein himself was very much concerned with attempts to constructing a unified field theory of gravity and electromagnetism. One might thus have expected that those attempts (and similar ones by Schrödinger and others) met with great interest at Berne. But this was not the case. Only Bruria Kaufman, Einstein's last collaborator, gave a main talk on the mathematical structure of the non-symmetric field theory, in which the Christoffel symbols Γ^s_{ik} are not required to be symmetric. The discussion after that talk contains only one mathematical comment from Marie-Antoinette Tonnelat.

The indifference towards Einstein's final attempts can well be understood. It had become obvious at least since the 1920s that quantum theory is needed to describe the atomic and subatomic world. The strong and weak interactions relevant for the microscopic regime are not taken into account in Einstein's work. Most physicists thus suspected (rightly) that an essential part of the world was missing in Einstein's attempts at a classical unification of gravity and electromagnetism. This opinion is clearly expressed in a letter of Pauli to Einstein from September 19, 1946:

> My personal opinion still is ... that the classical field theory in every form is a squeezed out lemon, out of which it is impossible to get anything new![9]

From a modern point of view, a more promising idea for a generalization was presented at the Berne meeting by Pascual Jordan. He gave an example of a theory with a 'varying gravitational constant'. Such a theory can be represented as a *scalar-tensor theory* of gravitation, in which a scalar field ϕ is added to the gravitational sector (see e.g. Fuji and Maeda 2003). The action for such "Jordan–Brans–Dicke theories" (as they were called later after the contributions by Brans and Dicke) reads

$$S_{\mathrm{JBD}} = \int \mathrm{d}^4x\sqrt{-g}\left(\phi R - \frac{\omega}{\phi}\phi_{,\mu}\phi_{,\nu}g^{\mu\nu} + \mathcal{L}_{\mathrm{m}}\right), \tag{4}$$

where ω is a new dimensionless parameter of the theory. Such theories (and generalizations thereof) are of much interest today, for example in connection with dark energy or the assumed inflationary phase of the early universe.

At the Berne conference, Jordan's contribution was received with scepticism. In his conference summary, Pauli 'buried' (an expression by Engelbert Schücking)[10] Jordan's theory as follows:

> By the magic of his mathematical theorems, Mr. Jordan has unfortunately prevented us from hearing something about his physical reasons to assume a variation of the gravitational constant; this would have surely interested all of us ...[11]

In our days, investigations into the variation of fundamental constants find general acceptance. The main reason for this is the expectation that a more fundamental theory than general relativity arises from the implementation of quantum theory (see below). In some of these theories, "constants" of Nature are described at high

---

[9]This is my translation from the original German which reads: "Meine persönliche Überzeugung ist nach wie vor ..., daß die klassische Feldtheorie in jeder Form eine völlig ausgepreßte Zitrone ist, aus der unmöglich noch etwas Neues herauskommen kann!" (von Meyenn 1993, p. 384).

[10]See (Harvey, 1999, p. 11).

[11]This is my translation from the original German which reads: "Nun hat uns leider Herr Jordan mit dem Zauber seiner mathematischen Sätze verhindert, etwas darüber zu hören, was eigentlich seine physikalischen Gründe sind, um eine Veränderung der Gravitationskonstante anzunehmen; das hätte uns ja sicher alle sehr interessiert ...". (Mercier and Kervaire 1956, p. 265).

energies by time- and space-dependent fields. Despite various searches, however, no time or space variation of "constants" was observed so far.

3 Cosmology

With the advent of Einstein's theory of general relativity, it was for the first time possible to provide a consistent description of the Universe as a whole. Assuming the cosmological principle, one arrives at the 'Robertson–Walker form' of the metric, from which the 'Friedmann–Lemaître equations' can be obtained from the Einstein equations. Today, one often speaks of 'Friedmann–Lemaître–Robertson–Walker' (FLRW) world models. On the observational side, not much was known at the time of the Berne Conference beyond Hubble's law and some crude age determinations. Still, cosmology was an important topic at the conference, certainly more important than one would have expected in retrospect. There were major reviews by Walter Baade (Mount Wilson Observatory) from the observational side and by Howard Robertson (California Institute of Technology) from the theoretical side. Baade did not deliver a manuscript to the Proceedings, so no statements can be made about his contribution. Robertson has sent a detailed manuscript that also contains a comparison of theory with observation.

It is not surprising that Robertson based his analysis on the homogeneous and isotropic Robertson–Walker metric. But he included the following important comment (p. 135 of the Proceedings):

> It is to be emphasized that we have not *required* the real universe to be one satisfying the uniformity conditions imposed above; we are merely examining the nature of the idealized model of the real world in which the obvious and all-important inhomogeneities are ironed out. We are not imposing the uniformity as a 'cosmological principle' …to which the real world must adhere.

In his contribution, Robertson discusses the observational status from a rather modern point of view. He presents a diagramme displaying the age of the universe against the matter density, and he allows for any value of curvature and cosmological constant Λ. The empirical value of the Hubble constant, H_0, at that time was given by 180 km/s Mpc, much higher than today's value.[12] The discrepancy of the historic value with today's value lies in the very crude distance measurements of the day, which have greatly improved since then.

Given the (too high) value of H_0 and conservative lower limits for the age of the Universe, Robertson finds that "…we are forced to reintroduce $\Lambda > 0$ in order to save this time scale …". Today we know that Λ (or its generalization in the form

[12]The current value from the Planck Collaboration (2018) is (67.4 ± 0.5) km/s Mpc, a bit more than one third of the 1955 value. There is currently a tension between cosmological and non-cosmological measurements of H_0. With the Hubble Space Telescope and the Gaia parallax measurements one gets (73.52 ± 1.62) km/s Mpc, see Riess et al. (2018).

of dark energy) is positive, although for a different reason than the one given by Robertson: it is because the Universe is found to be currently accelerating.

In addition to Baade's and Robertson's overviews, various shorter contributions on cosmology have been presented at the conference, including talks by Max von Laue, Oskar Klein, and Otto Heckmann. Heckmann, for example, presented a world model of Newtonian cosmology with expansion and rotation, which he had developed together with Engelbert Schücking. They had found that the introduction of rotation leads to a model without initial ('big bang') singularity. One might therefore wonder whether this can also happen in general relativity. This is, however, not the case. As the singularity theorems proven in the 1960s by Roger Penrose and Stephen Hawking show, singularities are unavoidable, given some general assumptions. But those theorems were not available at the time of the Berne Conference.

In 1955, the expansion of the Universe was not yet generally accepted. Ten years before the discovery of the Cosmic Microwave Background (CMB), it was still possible to seriously uphold the alternative model of a steady-state universe. Two of the main proponents of that model, Hermann Bondi and Fred Hoyle, were present in Berne and gave two short contributions. Today, this is of historic interest only.

Max Born, in his account of "Physics and Relativity", describing personal remembrances of the years around 1905, remarks that the importance of general relativity lies in the revolution which it has produced in cosmology. This is a remark that can certainly be appreciated today much more than in 1955.

4 Quantum Theory and Gravity

In the 1950s, quantum theory and its applications were at the centre of physics research worldwide. This is not surprising. In the realm of atoms, nuclei, and particles, plenty of new experimental results were found, and quantum theory, in its mechanical as well as field theoretical version, was believed to be the correct theory for their description. It is for this reason that the relation between quantum theory and relativity was discussed at length at the Berne Conference, too.

Eugene Wigner presented a major lecture on "Relativistic invariance of quantum-mechanical equations". He not only reviewed his important work on the representations of the Poincaré group but also discussed its extension to the de Sitter group. This is very interesting from a modern point of view because current observations indicate that our Universe is asymptotically approaching a de Sitter phase. Wigner emphasized that massive particles must then be characterized by the statement that their Compton wavelength is much smaller than c/H, where H is the (constant) Hubble parameter of de Sitter space. The relevance of de Sitter space for the formulation of asymptotic conditions is emphasized in Ashtekar et al. (2015).

A major problem is the consistent unification of quantum theory with gravity. This was open in 1955 and is still open today, in spite of much progress that has happened since then. Only two major talks were devoted to this problem,

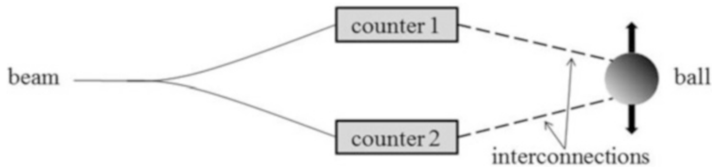


Fig. 4 Stern–Gerlach type of gedanken experiment, in which the detectors for spin up respective spin down are coupled to a macroscopic ball. If the particle has spin right, which corresponds to a superposition of spin up and down, the coupling leads to a superposition of the ball being moved up and down, leading to a superposition of the corresponding gravitational fields. Figure after DeWitt and Rickles, p. 251, see DeWitt (1957)

by Peter Bergmann and by Oskar Klein. Bergmann reviewed the state of the canonical formalism, which can serve as the starting point for the quantization of the gravitational field. This formalism was pioneered by Léon Rosenfeld in Zürich in 1930 and later developed in parallel by Bergmann and his group in Syracuse and by Paul Dirac in Cambridge as well as by Arnowitt, Deser, and Misner (ADM) in the United States.[13] Bergmann's talk was a bit dry in the sense that he restricted himself to pure formalism and did not address physical applications.

The real starting point of quantum gravity research is marked by a conference two years later. It took place at Chapel Hill, North Carolina, and was later known as the GR1 Conference. Many of the proponents of quantum gravity were present, including John Wheeler and Bryce DeWitt. Richard Feynman discussed there a gedanken experiment from which he concluded the necessity for quantizing the gravitational field, see Fig. 4.

Feynman concludes:[14]

> . . . if you believe in quantum mechanics up to any level then you have to believe in gravitational quantization in order to describe this experiment. . . . It may turn out, since we've never done an experiment at this level, that it's not possible . . . that there is something the matter with our quantum mechanics when we have too much *action* in the system, or too much mass—or something. But that is the only way I can see which would keep you from the necessity of quantizing the gravitational field. It's a way that I don't want to propose. . . . (DeWitt and Rickles, pp. 251–252, see DeWitt 1957)

The Berne Conference was still far behind this level of physical discussion. But in his concluding speech, Wolfgang Pauli expressed very clearly the main difficulty in quantizing the gravitational field. He said:

> This now leads to the border of knowledge, to the questions of the quantization of the field; it seems that a certain agreement existed in assuming that the mere application of conventional quantization methods probably will not lead to the goal. . . .

[13] See, for example, Kiefer (2012) for details.

[14] See also the discussion in Feynman et al. (1995).

> It seems to me ... that it is not so much the linearity or non-linearity which forms the heart of the matter [the difficulty of quantizing the gravitational field, C.K.], but the very fact that here a more general group than the Lorentz group is present.[15]

Standard ways of quantization assume the existence of a fixed background, which usually is taken to be Minkowski space. In general relativity, this background is absent—spacetime is dynamical, and the invariance is the diffeomorphism, not the Lorenz group. The quantization of the metric (which represents spacetime) has thus to be undertaken without any reference to Minkowski space with its Lorentz group; this is what Pauli is alluding to. Modern approaches to quantum gravity make use of this background independence (Kiefer, 2012).

In this connection, it is interesting to quote a piece from a letter that Pauli wrote to Schrödinger on the occasion of the latter's 70th birthday. In this letter, which is from August 9, 1957, he writes (von Meyenn, 2011, p. 720):[16]

> Also our difference in age of 13 years will soon appear as unessential, and one will count us as belonging to the same generation of physicists: to those who have e.g. not succeeded in making a synthesis of the mentioned subjects—general relativity and quantum theory—and who thus have left behind unsolved problems as essential as the atomistic nature of electricity (fine structure constant), self- energy of the electron ...[17]

In his response from August 15, 1957, Schrödinger writes (von Meyenn, 2011, p. 722):

> You are, of course, right, that we belong to the same generation of physicists; I also agree with your characterization of it. But the posterity usually judges in a milder way, it characterizes an epoch by what it has achieved, much more rarely by what it has not completed.[18]

Even today, the problem of quantum gravity remains unsolved. The main approaches, more or less promising, are direct quantizations of general relativity in either its canonical or covariant form and string theory. The latter is characterized

---

[15]This is my translation from the original German which reads: "Das führt nun hier an die Grenze des Wissens, an die Fragen der Quantisierung des Feldes; es scheint, daß eine gewisse Übereinstimmung darüber bestand, daß eine bloße Anwendung konventioneller Quantisierungsmethoden wahrscheinlich nicht zum Ziele führen wird. ...

Es scheint mir ..., daß nicht so sehr die Linearität oder Nichtlinearität Kern der Sache ist, sondern eben der Umstand, daß hier eine allgemeinere Gruppe als die Lorentzgruppe vorhanden ist." (Mercier and Kervaire 1956, p. 266).

[16]I thank Norbert Straumann for drawing my attention on this and the following letter.

[17]This is my translation from the original German which reads: "Auch unser Altersunterschied von 13 Jahren wird bald als unwesentlich erscheinen, und man wird uns zur selben Physiker-Generation zählen: zu derjenigen, der z.B. eine Synthese der beiden genannten Themen – allgemeine Relativitätstheorie und Quantentheorie – nicht gelungen ist und die so wesentliche Probleme wie Atomistik der Elektrizität (Feinstrukturkonstante), Selbstenergie des Elektrons ... ungelöst zurückließ."

[18]This is my translation from the original German, which reads: "Natürlich hast Du recht, daß wir zu derselben Physikergeneration gehören; auch dem, wie Du sie kennzeichnest, stimme ich bei. Nur pflegt die Nachwelt milder zu sein, sie pflegt eine Epoche zu charakterisieren nach dem, was sie geleistet hat, viel seltener nach dem, was sie nicht fertig gebracht hat."

by an attempt to construct, at a fundamental level, a unified quantum theory of all interactions (sometimes called 'theory of everything'), from which quantum gravity can be recovered in a certain limit. A central problem for all attempts is the current lack of experimental support. The only exception is an indirect test of linearized quantum gravity: adopting the inflationary scenario of the early Universe, the power spectrum of the CMB is proportional to the Planck time squared and needs the quantization of the metric for its calculation.[19] The density fluctuations in the CMB have been observed and are in accordance with this prediction. The influence of primordial gravitons has not been seen yet, but this is in principle possible; its observation would be a clear test of (linearized) quantum gravity.

Oskar Klein's talk at Berne was of a more general nature. Like many other physicists at the time, he was worried about the divergences in quantum field theory. In his proposed generalization of general relativity, he went beyond Einstein's own attempts (which he didn't cite) and discussed the five-dimensional theory which today is known as Kaluza–Klein theory (but he does not cite Kaluza here). For him, this theory is the most direct generalization of Einstein's theory including gauge invariance and charge conservation. As a motivation, Klein directly referred to nuclear and mesonic physics, for which this theory should be relevant. He attributed a fundamental role to the five-dimensional Dirac equation in the sense that it is prior to geometry: the components of the Riemann tensor follow from the commutator of the covariant derivatives

$$\Delta_\mu \psi \equiv \left(\frac{\partial}{\partial x^\mu} - \Gamma_\mu \right) \psi, \tag{5}$$

where Γ_μ denotes the connection, and ψ is the Dirac spinor. Gravity is supposed to play an important role when kinetic energies approach the Planck scale, and Klein speculated that gravity may serve as a natural regulator for the field theoretical divergences. In this, he directly related the compactification radius of the five-dimensional theory to the Planck length.

Higher dimensions play an important role in string theory, which is probably consistent only in ten spacetime dimensions. The theory of supergravity is consistent in any dimension up to eleven; supergravity may play an important role in the speculative M-theory. In string or in supergravity theory, as well as in direct quantizations of general relativity, gravity may directly or indirectly serve as a regulator for the field theoretic divergences, although the final word has not been spoken yet.

[19]The Planck time is $t_{\rm P} \equiv \frac{l_{\rm P}}{c} = \sqrt{\frac{\hbar G}{c^5}} \approx 5.39 \times 10^{-44}$ s ($l_{\rm P}$ is the Planck length.)

5 Summary

One can state that the Berne Conference of 1955 marks a turning point in the history of relativity. This fact is also emphasized in Blum et al. (2015). It was the first truly international conference on general relativity and its generalizations. The importance and prospects of those theories for the future are reflected in many contributions to the Proceedings. Today, Einstein's theory is empirically well tested, and the fields of cosmology and quantum gravity occupy a central place in current research.

Acknowledgments I am grateful to the organizers of the conference *Thinking about Space and Time* for inviting me to such an inspiring event. I thank Cormac O'Raifeartaigh and Norbert Straumann for their comments on my manuscript and Stanley Deser for sharing his memories of the Berne Conference.

References

Ashtekar, A., Bonga, B., and Kesavan, A., 2015. Asymptotics with a positive cosmological constant: I. Basic framework. *Classical and Quantum Gravity* **32**, 025004 (41 pp).

Ashtekar, A. and Petkov, V. (eds.), 2014. *Springer Handbook of Spacetime*. Dordrecht: Springer.

Blum, A., Lalli, R., and Renn, J., 2015. The Reinvention of General Relativity: A Historiographical Framework for Assessing One Hundred Years of Curved Space-time. *Isis* **106**, 598–620.

Bruns, D. G. (2018). Gravitational starlight deflection measurements during the 21 August 2017 total solar eclipse. *Classical and Quantum Gravity* **35**, 075009 (21pp).

Chruściel, P. T. and Friedrich, H. (eds.), 2004. *The Einstein Equations and the Large Scale Behavior of Gravitational Fields* (Birkhäuser, Basel).

DeWitt, C. (ed.), 1957. *Proceedings of the conference on the role of gravitation in physics*, University of North Carolina, Chapel Hill, January 18–23, 1957. WADC Technical Report 57–216 (unpublished). These Proceedings have recently been edited in: D. Rickles and C. M. DeWitt (eds), Edition Open Sources, http://www.edition-open-sources.org/sources/5/

Eckart, A. et al., 2017. The Milky Way's Supermassive Black Hole: How Good a Case Is It? *Foundations of Physics* **47**, 553–624.

Feynman, R. P., Morinigo, and F. B., Wagner, W. G., 1995. *Feynman Lectures on Gravitation*. Reading: Addison-Wesley.

Fölsing, A., 1994. *Albert Einstein*. 3rd ed. Frankfurt am Main: Suhrkamp.

Fuji, Y. and Maeda, K, 2003. *The scalar-tensor theory of gravitation*. Cambridge: Cambridge University Press.

Harvey, A. (ed.), 1999. *On Einstein's Path*. New York: Springer.

Isenberg, I., 2014. In: Ashtekar and Petkov (2014), pp. 303–321.

Kiefer, C., 2012. *Quantum Gravity*. 3rd ed. Oxford: Oxford University Press.

Mercier, A. and Kervaire, M. (eds.), 1956. *Helvetica Physica Acta Supplementum IV* (Birkhäuser, Basel). This volume is available online on https://www.e-periodica.ch.

Planck Collaboration, 2018. Planck 2018 results. VI. Cosmological parameters. arXiv:1807.06209 [astro-ph.CO].

Riess, A. G. et al., 2018. Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant. *Astrophysical Journal* **861**, 126 (13pp).

von Meyenn, K. (ed.), 1993. *Wolfgang Pauli. Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a.* Band III. Berlin: Springer.
von Meyenn, K. (ed.), 2011. *Eine Entdeckung von ganz außerordentlicher Tragweite*. *Schrödingers Briefwechsel zur Wellenmechanik und zum Katzenparadoxon*. Berlin: Springer.
Will, C. M., 2014. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity*, available online as https://doi.org/10.12942/lrr-2014-4.

Einstein's Conflicting Heuristics: The Discovery of General Relativity

John D. Norton

Abstract Einstein located the foundations of general relativity in simple and vivid physical principles: the principle of equivalence, an extended principle of relativity and Mach's principle. While these ideas played an important heuristic role in Einstein's thinking, they provide a dubious logical foundation for his final theory. Einstein was also guided to his final theory, I argue, by a second tier of more prosaic heuristics. I trace one strand among them. The principle of equivalence guided Einstein well until it led him to a theory that contradicted the conservation of momentum. Einstein converted the requirement of conservation of energy and momentum into a procedure that he used repeatedly for finding gravitational field equations. That procedure survives in present day developments of general relativity.

1 Introduction

What were the heuristics that guided Einstein to his completed general theory of relativity of 1915? There can be no simple answer. The completion of the theory came only after eight years of exhausting labor. In them, Einstein, at the height of his creative powers, grappled with problems so profound that they nearly defeated him. Nonetheless, Einstein himself provided an appealing and simple narrative of his discovery. He was guided, he assured us, by a few simple but powerful physical principles and thought experiments. These same heuristics then became the basis of Einstein's later account of the logical foundations of general relativity.

In narrowing his focus to these few heuristics, Einstein purged his account of nearly all the complications and false steps that later historical work has revealed. It obscures the fact that there is a great distance between the lofty generalities of Einstein's principles and the messy details of the final theory. These principles could not by themselves have led Einstein to the final theory. Worse, as will be recounted

J. D. Norton (✉)
Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: jdnorton@pitt.edu; http://www.pitt.edu/~jdnorton

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_2

in Sect. 2 below, most of the heuristics of this narrative turn out not to be vindicated by Einstein's final theory and may even fail to be sustainable as independent ideas. They provided at best an unreliable guide and a dubious logical foundation for the theory.

So we must ask again, what were the heuristics that guided Einstein? Something more must have helped Einstein arrive at his final success. My contention in this paper is that beneath this first tier of heuristics lies a second tier of heuristics. They do not lend themselves to arresting statements of a grand vision. Rather they are the practical lessons that a theorist like Einstein learns as he returns day after day to the mundane work of theory building. Whatever else may happen, Einstein's new theory must conserve energy and mesh with Newton's old theory of gravity. Getting all these details to work is not easy. A theorist can readily be led into blind alleys. The theorist must learn the tricks that avoid the traps. Once they are learned, it is all too easy to omit them from the celebratory recollections. However it is the accretion of these lesser heuristics that proves as important to the final discovery. Without them, the final result could not be achieved.

This paper traces the fortunes of just one of Einstein's first tier heuristics, the principle of equivalence. It did guide Einstein's thinking. However the principle was defeasible. We shall see that it was diluted in 1912 and all but discarded in 1913 when a second tier heuristic, the requirement of conservation of energy and momentum, led to gravitational field equations that contradicted it. This second tier heuristic was beyond challenge. It persisted and powerfully circumscribed Einstein's continuing analysis up to the completion of his theory in 1915.

In the following, Sects. 2, 3, 4, 5, and 6 describe the origin of the principle in Einstein's earliest reflections on gravitation and acceleration and traces how it guided Einstein to a novel theory of static gravitational fields in 1912. Sections 7, 8, 9, 10, and 11 recall how Einstein found that the resulting theory conflicted with a second tier heuristic, the conservation of momentum. Reluctantly, Einstein was compelled to modify the theory's single field equation to one that compromised his principle of equivalence. The principle could now only hold, as he put it, for infinitely small fields.

The following year, as recounted in Sect. 12, Einstein and his mathematician friend Marcel Grossmann devised the "Entwurf" theory. It differed from general relativity only in employing gravitational field equations of limited covariance. Conservation of momentum had, in 1912, forced a quite specific field equation on Einstein. He now turned that experience to his advantage. Einstein went to pains to explain in elementary terms that the conservation requirement provides a general method for arriving at unique field equations. He then used it with the conservation of energy and momentum to identify the gravitational field equations of "Entwurf" theory. What Einstein did not then acknowledge clearly, as Sect. 13 recalls, was that his original principle of equivalence now failed completely with this choice of field equations. Whatever the merits of this first tier heuristic, its role in theory formation was quite displaced by the second tier heuristic of the conservation of energy and momentum.

This second tier heuristic enjoyed a brief moment of prominence when it was highlighted in Einstein and Grossmann's (1913) "Entwurf" paper as the foundation of the method used to derive the theory's gravitational field equations. Section 14 recounts how the heuristic became less visible but continued to exercise a controlling role in Einstein's subsequent theorizing. It governed Einstein's analysis of the limited covariance of the "Entwurf" theory and persists in modern accounts of general relativity in providing the fastest route to Einstein's celebrated gravitational field equations of 1915.

Section 15 and the concluding Sect. 16 review the tension between the conception and application of the heuristics of the two tiers. Einstein's recounting accords the first tier heuristic, the principle of equivalence, primary foundational importance. Yet it was defeasible in his actual theorizing when it conflicted with a second tier heuristic, the conservation of energy and momentum. Section 15 also reviews briefly a related episode of heuristics in tension that was explored in some detail by Janssen and Renn (2007). It concerns Einstein's November 1915 return to generally covariant field equations. Two appendices contain background calculations.

2 Einstein's Principles

Einstein completed his general theory of relativity in November 1915. The triumph came to an exhausted and exhilarated Einstein after 8 years of labor on the problem of relativity and gravitation. It was a distinctive achievement, quite unlike so many other discoveries in physics. In these other cases, novel empirical results were key. The final theory lay hidden in them in encoded form. Success came when someone figured out how to read the code. The nineteenth century accumulated a wealth of empirical results on electricity and magnetism. They were summarized in the Maxwell–Lorentz electrodynamics that the young Einstein studied so eagerly. Encoded in them he found the Lorentz transformation and with it the special theory of relativity of 1905.[1] In the same year, Einstein found his revolutionary light quantum hypothesis. It was encoded, he realized, in the recently measured thermodynamic properties of heat radiation.[2]

The discovery of general relativity was quite unlike these. There was some empirical guidance. Perhaps the most significant empirical result guiding Einstein was an old one. It was a commonplace since the work of Galileo and Newton in the seventeenth century that all bodies fall under gravity with the same acceleration, independently of their masses. Aside from this result, on his own account, the heuristics that guided Einstein were more ethereal and philosophical in character. At their center was what Einstein labeled "an epistemological defect" in prior

[1] See Norton (2004) for an account of Einstein's investigations.

[2] See Norton (2006) for an account of this encoding.

theories of both classical mechanics and special relativity.[3] These theories were defective in positing inertial frames of reference since their disposition was fixed absolutely without relation to the contents of space and time. The associated, preferred inertial motions were absolute in a sense Einstein (1923, p. 61) found objectionable: “independent in its physical properties, having a physical effect, but not itself influenced by physical conditions.” To eliminate this defect, Einstein proposed that the principle of relativity had to be extended from the relativity of inertial motion of his 1905 special theory of relativity to include accelerated motion as well.

The need for this extension was grounded further in an idea that Einstein attributed[4] to Ernst Mach: that the inertia of bodies is due to an interaction with the other masses of the universe. This, he labeled “Mach’s principle.” According to it, the distribution of matter in space determines completely the disposition of the inertial frames of reference. Finally there was what Einstein called[5] the “happiest thought of [his] life,” the principle of equivalence. It asserted the equivalence of uniform acceleration in gravitation free space and a homogeneous gravitational field. This principle, Einstein was already able to boast at the outset in (Einstein 1907, p. 454), “extends the principle of relativity to uniformly accelerated translational motion of the reference system.” It was, he felt, a promising first step.

These heuristics are widely celebrated. They are almost as well-known as the iconic photographs of Einstein, the disheveled genius and iconoclast. Their popularity is driven by their vividness and simplicity. They lend themselves to memorable thought experiments. The principle of equivalence is routinely expressed through the parable of an observer trapped in a box or an elevator. The box is accelerated in gravitation free space; or, in later variants, the box is in free fall in a gravitational field. Mach’s principle is routinely related as an answer to Newton’s own bucket thought experiment. Newton had proposed in his *Principia* that the concavity in the surface of the water in a rotating bucket arises from acceleration with respect to absolute space. Mach’s principle asserts that, instead, the concavity arose from the water’s acceleration with respect to all the other masses of the universe.

These heuristics promise an easy pathway to understanding a theory that, reputedly, is so abstruse that few can properly understand it.[6] Everyone who has driven in a car understands viscerally how acceleration produces inertial forces. These, we are told on Einstein’s authority, are just the same thing as gravitational forces; and they arise precisely because you are accelerating in relation to all the

---

[3]“ein erkenntnistheoretischer Mangel” Einstein (1916, p. 771).

[4]There is some question over whether Einstein’s attribution to Mach was correct. See Norton (1995).

[5]Einstein (1920a, p. 265).

[6]An anonymous preface to Lorentz (1920, p. 5) Begins “Whether it is true or not that more than twelve persons in all the world are able to understand Einstein’s Theory, . . . ”

other masses of the universe. Appreciate that and an understanding the general relativity of all motion is almost within your grasp. It seems so easy.

Einstein's autobiographical statements leave no doubt of the importance of these heuristics in Einstein's process of discovery. However there is a troubling aspect to them. They depend heavily on judgments of how physical theories have to be, independent of experience. Such efforts are rarely successful. Time and again today's experience or more careful thought has overturned yesterday's theoretical imperatives. So it is with Einstein's heuristics. Many do not survive scrutiny.[7]

There is no deeper principle of nature that requires us to eschew something that has (in Einstein's words) "a physical effect, but [is] not itself influenced by physical conditions." Whether inertial frames of reference are as special relativity dictates is a matter to be decided empirically and not by a priori stipulation. Contrary to Einstein's earlier hopes, the Machian principle turned out not to be implemented in the final general theory of relativity and he eventually abandoned the principle. The generalization of the principle of relativity to accelerated motion was implemented by Einstein as a demand that his new theory be expressible in arbitrarily chosen space-time coordinate systems. Kretschmann quite correctly objected in 1917 that this requirement was all but vacuous. It was more a challenge to the ingenuity of theorists in the way they wrote their equations. Finally Einstein's original formulation of the principle of equivalence almost immediately disappeared from the literature. In its place came a proliferation of variant forms ("weak," "strong," "Einstein") that differed from Einstein's in both fundamental conception and content.

Troubled as these heuristics are, there is no doubt of their importance in Einstein's mind while he worked on the problem of relativity and gravitation. If they were his only guides, then it would be somewhat more than extraordinary that his deliberations should produce such a remarkable result, the general theory of relativity. There were, as we shall now see, many more guides. They were buried in details that did not lend themselves to popular exposition.

The attempt in this paper to understand how Einstein succeeded nonetheless proceeds in the spirit of Janssen (2014), who addresses the same question. In his concluding Sect. 6, entitled "Post Mortem: How Einstein's Physics Kept his Philosophy in Check," Janssen attributes Einstein's success to three factors:

> First, Einstein did not just want to eliminate absolute motion, he also wanted to reconcile some fundamental insights about gravity with the results of special relativity and integrate them in a new broader framework. Second, when these efforts led him to the introduction of the metric field, he carefully modeled its theory on the successful theory of the electromagnetic field of Maxwell and Lorentz. Third, whenever his philosophical agenda clashed with sound physical principles, Einstein jettisoned parts of the former instead of compromising the latter.

The analysis of this paper illustrates the third of these factors. Einstein's principle of equivalence belongs in what Janssen calls the "philosophical agenda." It is here

[7]For a synoptic survey of the problems in Einstein's heuristics, see Norton (1993).

a defeasible, first tier heuristic. Conservation of energy and momentum is one of Janssen's "sound physical principles." It belongs in the second tier of heuristics that cannot be compromised.

3 Einstein's 1907 Heuristic

The project began in 1907 when Einstein was commissioned to write a review article on the "principle of relativity," this being the term that delineates what we would now call the special theory of relativity. The resulting review article, Einstein (1907), showed how existing branches of physics could accommodate or be accommodated to Einstein's new theory of space and time. Only one area of physics proved troublesome: gravitation. In Sect. 5, Einstein embarked on a speculative new approach to gravity that might at the same time afford an extension of the principle of relativity to accelerated motion.

The heuristic device that guided Einstein was labeled merely as an "assumption" (*Annahme*). In what we must presume was the space of Newtonian mechanics, he considered a reference system Σ_1 uniformly accelerated in a fixed direction and a second inertial reference system Σ_2 in which there is a homogeneous gravitational field. He supposed further that the acceleration of Σ_1 matched the acceleration of fall of free bodies in Σ_2, so that the motions of free bodies would be the same in both systems. Einstein's assumption was that this sameness was to be generalized to all physical processes. We must presume a tacit extension to relativistic contexts. He wrote (p. 454):

> We have therefore in the present state of our experience no basis for the assumption that the systems Σ_1 and Σ_2 differ from one another in any respect. Hence we want to assume in the following the complete physical equivalence of a gravitational field and the corresponding acceleration of the reference system.
>
> This assumption extends the principle of relativity to the case of uniformly accelerated translational motion of the reference system . . . .

Modern readers will immediately recognize this as Einstein's first statement of the principle of equivalence. They may however be puzzled by the restriction of equivalence to the special case of a homogeneous gravitation field and uniform acceleration. Standard modern statements of the principle of equivalence are more general. They commonly assert that a gravitational field can always be transformed away, at least locally, by adopting an appropriate acceleration of the reference system.

That this more general version of the principle was not Einstein's has been recounted in Norton (1985). We need not rehearse here Einstein's objections to the generalized principle. The important point is to recognize that the principle was, for Einstein in 1907, not yet a permanent axiom of some well-articulated theory. That may still come. In 1907, the primary interest of the assumption for Einstein was as a heuristic guide in the generation of a new theory of gravity, whose general outlines were only dimly visible to Einstein in 1907. Einstein stated clearly his heuristic purpose in the continuation of the passage quoted above

> ... The heuristic value of the assumption lies in the fact that a homogeneous gravitational field may be replaced by a uniformly accelerated reference system. The latter case is accessible to theoretical treatment to a certain degree.

There was then, in Einstein's view, an urgent need for such a heuristic. For Einstein had tried an obvious accommodation of gravity to special relativity, that is, the construction of simple, Lorentz covariant theories of gravity. Einstein (1933, pp. 286–287) recalled the problem he discovered:

> These investigations, however, led to a result which raised my strong suspicions. According to classical mechanics, the vertical acceleration of a body in the vertical gravitational field is independent of the horizontal component of its velocity. Hence in such a gravitational field the vertical acceleration of a mechanical system or of its center of gravity works out independently of its internal kinetic energy. But in the theory I advanced, the acceleration of a falling body was not independent of its horizontal velocity or the internal energy of the system.
>
> This did not fit in with the old experimental fact that all bodies have the same acceleration in a gravitational field ...

In short, Einstein had failed to find a relativized theory of gravity in which bodies fall vertically with equal acceleration, independently of their horizontal motion.[8] How could Einstein proceed? The assumption of 1907—the principle of equivalence—provided a way. It delivered to Einstein a single instance of a gravitational field with just the independence property needed. To proceed, all Einstein needed to do was to catalog the properties of this one special case of a relativized gravitational field and then judiciously generalize them to recover a full theory.

4 Einstein 1907–1912 Theory of Static Gravitational Fields

This project of generalization became the substance of those parts of the ensuing five years that Einstein devoted to gravitation. The 1907 review article already contained some now familiar results. The speed of light and the ticking of clocks would be slowed in a homogeneous gravitational field. This speed played the role of a gravitational potential. These results, now generalized to the inhomogeneous static gravitational field of the sun, yielded a prediction of a slight red shift in light emitted by the sun. Einstein returned to work on the theory in 1911, when he realized that another effect in it was open to observational test. According to the theory (Einstein 1911), a beam of light is bent by a gravitational field. The bending should be detectible in a displacement of apparent star positions in the sky in the vicinity of the sun.

[8] For an attempted reconstruction of Einstein's explorations, see Norton (1992, §3).

While this 1911 analysis is widely known through its inclusion in the ubiquitous Dover reprint *The Principle of Relativity*, the fullest expression of the project of generalization came in a lesser-known pair of papers the following year (Einstein 1912a, b). These papers contained a full theory of certain static gravitational fields. The theory provided equations of motion for bodies in free fall, a field equation for the variable speed of light and versions of electrodynamics and thermodynamics, suitably modified to accommodate the novelty of a variable speed of light.

The starting point of the paper is a transformation from the familiar reference system Σ of Einstein's 1905 special theory of relativity, represented by coordinates of space and time (ξ, η, ζ, τ), to a unidirectionally, uniformly accelerated frame of reference K, represented by the coordinates of space and time (x, y, z, t). Einstein's analysis is cumbersome. He does not develop the full transformation equations, although (as we are about to see), they are quite simple. In a labored development proceeding over many pages, he recovers only an approximation of the general transformation equations for small t. Einstein's generalizations proceed from them.

Here I will not recapitulate these details. They would provide no special illumination for the issues to be raised. Instead I will summarize them using a more perspicacious formalism that Einstein himself shortly recognized. In a last minute correction to the proofs of Einstein (1912b, p. 458), Einstein found his equations of motion is recovered most simply from an action principle. The following year Einstein and Grossmann (1913, p. 7) revealed that this action principle is the equation of geodesic motion in a spacetime whose structure is no longer Minkowskian.

5 The Gravitational Field of Uniform Acceleration

The equations relating the unaccelerated and uniformly accelerated frame Σ and K were given later in many places, including Einstein and Rosen (1935, p. 74):[9]

$$\begin{aligned} \tau &= (c_0/a + x)\ \sinh(at) \\ (\xi + c_0/a) &= (x + c_0/a)\ \cosh(at) \qquad \eta = y \qquad \zeta = z \end{aligned} \tag{1}$$

The acceleration is uniform translational acceleration in the ξ, x direction; a is a constant acceleration parameter; and c_0 a constant. While the original coordinates of Σ cover the whole of the spacetime, those of K cover only a wedge delimited by null surfaces $\tau = (\xi + c_0/a)$ and$\tau = -(\xi + c_0/a)$. The x coordinate can only take values greater than $-c_0/a$, for the coordinates are singular at $x = -c_0/a$, where all

[9]The notation is adapted to Einstein's (1912a) usage. The Einstein and Rosen version was the slightly simpler case in which c_0=0.

the hypersurfaces of constant t intersect.[10] When t is small, the hyperbolic functions in (1) are well-approximated as $\sinh(at) \approx at$ and $\cosh(at) \approx 1 + a^2t^2/2$. Then the exact transformation equations (1) are well-approximated by the small t expressions Einstein derived in Einstein 1912a, p. 359):

$$\tau = (c_0 + \mathrm{a}x)\, t \qquad \xi = x + (c_0 + \mathrm{a}x)\, at^2/2 \qquad \eta = y \qquad \zeta = z \tag{2}$$

Under the transformation (1), using the perspectives Einstein would develop the following year, the Minkowskian expression for the invariant line element[11]

$$ds^2 = d\tau^2 - d\xi^2 - d\eta^2 – d\zeta^2 \tag{3}$$

becomes

$$ds^2 = (c_0 + ax)^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{4}$$

While the transition from expression (3) to (4) has merely redistributed the coordinates assigned to events, Einstein used the principle of equivalence to conclude that a homogeneous gravitational field now manifests in the new frame of reference $K(x, y, z, t)$. We can read directly from the line element (4) the same properties that Einstein inferred for this field, but with greater effort on his part.

Einstein defined the speed of light c in terms of the new coordinates as

$$c^2 = (dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2 \tag{5}$$

where $(x(t), y(t), z(t))$ is the trajectory of a light pulse. We read immediately from (4) that this speed of light c varies linearly with x in the direction of the gravitational field[12]

$$c\,(x, y, z) = c_0 + ax \tag{6}$$

Hence it can represent the gravitational potential.

We also read from (4) that the hypersurfaces of constant t are ordinary Euclidean spaces and that their coordinates (x, y, z) are Cartesian coordinates with the familiar metrical significance. The same is no longer true of the time coordinate t. It can no longer be measured directly by clocks. Rather times elapsed on a clock at rest in the

[10] In 1912, since he worked only with a small t approximation (2), Einstein may not have realized that the coordinates (x, y, z, t) he introduced have a singularity at $x = -c_0/a$. Einstein and Rosen (1935) later suggest that one can conceive the "field-producing mass" as located at this singularity, although they seek to eliminate the singularity. In (1912a, p. 356, footnote), however, Einstein wrote: "The masses that produce this field should be conceived as at infinity."

[11] The notation is adapted to Einstein's (1912a) usage. The Einstein and Rosen version was the slightly simpler case in which $c_0=0$.

[12] The other case of $c(x, y, z) = -(c_0 + ax)$ is not mentioned by Einstein.

frame must be rescaled by a position dependent factor $(c_0 + ax)$ if the corresponding time coordinate t differences are sought.

The equations of motion of bodies in free fall in this homogenous field are recovered by seeking the geodesics of the spacetime, that is, those trajectories for which $\int ds$ is extremal. A short and standard calculation of the Euler–Lagrange equations yields

$$\frac{d}{dt}\left(\frac{\dot{x}/\,(c_0 + ax)}{\sqrt{1 - q^2/(c_0 + ax)^2}}\right) = \frac{-a}{\sqrt{1 - q^2/(c_0 + ax)^2}}$$

$$\frac{d}{dt}\left(\frac{\dot{y}/\,(c_0 + ax)}{\sqrt{1 - q^2/(c_0 + ax)^2}}\right) = \frac{d}{dt}\left(\frac{\dot{z}/\,(c_0 + ax)}{\sqrt{1 - q^2/(c_0 + ax)^2}}\right) = 0 \qquad (7)$$

where $q^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$ and the overhead dot denotes differentiation with respect to t, $\dot{x} = dx/dt$, etc.

6 Einstein Generalizes Naturally

With these results in hand for the special case of a homogeneous gravitational field, Einstein could now proceed with his project of generalization. The generalizations he introduced were obvious and natural.[13]

First, the line element (4) is replaced by the more general

$$ds^2 = c^2\,(x, y, z)\,dt^2 - dx^2 - dy^2 - dz^2 \qquad (8)$$

where c can now vary more generally as a function of the spatial coordinates. This variable speed of light c still serves as the single gravitational potential and the spatial hypersurfaces of constant time coordinate t remain Euclidean. This new structure represents a more general case of time independent gravitational fields. Einstein recognized explicitly (1912a, p. 356), however, that it was not the most general case. He noted that the field produced by a rotation of the reference frame would yield a non-Euclidean geometry. For the Lorentz contraction would act differentially on rods oriented parallel or transverse to the direction of rotation. That

[13]This project may seem familiar since it is the first instance of what becomes the gauge argument routinely used to introduce interacting fields in particle physics. One starts with a flat connection, the case of no interaction. It is re-coordinatized (or its gauge changed) so that the new description mimics an interaction, while none is actually present. The new description is generalized to return equations governing a non-trivial interaction.

meant that the ratio of the circumference of a suitably placed circle to its diameter would no longer be the Euclidean value of π, when both are measured by rods at rest in the rotating reference frame.

In the generalization, the gravitational field strength is the negative gradient of the speed of light, ($-\partial c/\partial x$, $-\partial c/\partial y$, $-\partial c/\partial z$). The equations of motion of a body in free fall in a homogeneous gravitational field (7) are naturally generalized to

$$\frac{d}{dt}\left(\frac{\dot{x}/c}{\sqrt{1-q^2/c^2}}\right) = \frac{-\partial c/\partial x}{\sqrt{1-q^2/c^2}}$$

$$\frac{d}{dt}\left(\frac{\dot{y}/c}{\sqrt{1-q^2/c^2}}\right) = \frac{-\partial c/\partial y}{\sqrt{1-q^2/c^2}} \quad \frac{d}{dt}\left(\frac{\dot{z}/c}{\sqrt{1-q^2/c^2}}\right) = \frac{-\partial c/\partial z}{\sqrt{1-q^2/c^2}} \tag{9}$$

These equations coincide with the geodesics of the line element (8). They can also be recovered directly by solving the Euler–Lagrange equations using this more general line element (8).

Finally, Einstein sought a more general equation governing the speed of light c. The linear dependence of c on x in (6) is easily seen to be a solution of the Laplace equation for c:

$$\Delta c = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) c = 0 \tag{10}$$

In turn, it is naturally generalized to:

$$\Delta c = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) c = kc\sigma \tag{11}$$

for k a constant and σ the matter density. This field equation is the obvious analog of Poisson's equation for the Newtonian gravitational potential φ

$$\Delta\varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \varphi = 4\pi k\rho \tag{12}$$

for mass density ρ (in the form given in Einstein and Grossmann 1913, p. 11).

The principal difference in form between (11) and (12) is that the first has a source term $kc\sigma$ that is linear in the potential c, whereas the source term of the second, $4\pi k\rho$, has no corresponding term in φ. This difference reflects a difference in gauge freedoms in the two quantities c and φ. The speed of light c is undetermined up to a multiplicative factor M, reflecting our freedom to choose measuring units for distances and times. Thus, if c is a solution of (11) for some σ, then so is $c' = cM$. The Newtonian potential φ, however, is undetermined up to an additive factor A. Thus, if φ is a solution of Eq. (12) for some ρ, then so is $\varphi' = \varphi + A$.

This difference in gauge freedoms in the two cases may now seem innocuous. It will shortly prove to be a cause of considerable trouble.

7 A Hidden Peril

The generalizations of the last section are small and modest. They would be, it seems, just a small and secure step towards the most general theory. However as Einstein would shortly discover, these generalizations were far from innocent. The danger lay precisely in their apparent modesty, so that one would not readily think to challenge them.

Buried in the generalizations were two, specific problems. The first was the idea that space would remain Euclidean in the case of more general static gravitational fields. This proves almost never to be the case. Take one of the simplest cases: the Schwarzschild spacetime, the exterior gravitational field of a rotationally symmetric, uncharged, non-rotating body of mass m. Its line element is

$$ds^2 = \left(1 - \frac{2Gm}{r}\right) dt^2 - \frac{dr^2}{\left(1 - \frac{2Gm}{r}\right)} - r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{13}$$

The constant G is the Newtonian universal constant of gravitation and (r, θ, ϕ) are spherical coordinates of space. The failure of Euclidean geometry for the spatial hypersurfaces of constant t arises through the division of dr^2 by the factor $(1-2Gm/r)$. For, without it, the spatial line element is the Euclidean $dr^2 + r^2 (d\theta^2 + \sin^2\theta \, d\phi^2)$.

The trouble with Einstein's 1912 assumption of spatial flatness is that it is incompatible with his final field equations of November 1915. As Stachel (1989) first pointed out, as long as Einstein expected fields like (8) to satisfy his field equations, he is precluding the source free field equations of the vanishing of the Ricci tensor, $R_{\mu\nu} = 0$.[14] When Einstein adopted the mathematical framework of general relativity with his joint work with Grossmann (Einstein and Grossman, 1913), notoriously, Einstein considered and rejected generally covariant gravitational field equations employing the Ricci tensor. This misstep marked the beginning of years of painful drifting, while Einstein sought to reconcile himself with a misshapen theory. Those years brought Einstein's formulation of his "hole argument." It sought to establish that generally covariant gravitational field equations would not be physically interesting. The assumption of spatial flatness supported his earlier prediction of only a "half deflection" in a beam of starlight grazing the sun. Einstein (1915b) found this error only at the last moment in November 1915, when his

[14]The precise result is shown below in Appendix 1. If the metric tensor is restricted to the form (36), then the vanishing of the Ricci tensor permits the $g_{00} = c^2(x, y, z)$ to vary at most linearly with the coordinates x, y, and z as in (42).

celebrated computation of Mercury's anomalous motion depended on the failure of Euclidean geometry in the vicinity of the sun.

This one mistaken assumption was not the sole source of these years of misery for Einstein. However it was their starting point. I need only here reaffirm the profound and extended misery this assumption visited upon Einstein. For this episode has been the subject of very extensive historical investigations elsewhere, to which the reader is now directed. See Stachel (1989); Norton (1984), and for a synoptic work by Michel Janssen, John D. Norton, Jürgen Renn, Tilman Sauer, and John Stachel that significantly develops these earlier papers, see Renn (2007).

8 A Second Hidden Peril Identified

While this last peril lingered on unnoticed for several years, there was a second peril that Einstein identified almost immediately. Einstein's first paper of 1912 (Einstein 1912a) had been submitted to *Annalen der Physik* on February 26, 2012. Before its printing was finalized, Einstein found to his dismay that the natural and obvious field Eqs. (10) and (11) could not be exactly correct. He managed to append a footnote (p. 360) to them that alerted readers to the problem:

> In a work to follow shortly, it will be shown that the Eqs. (10) and (11) still cannot be exactly correct. They will be used provisionally in this work.

The work promised, Einstein (1912b), was submitted on March 23, 2012, to the journal, just under a month after the first paper was submitted. It dealt first with routine matters required by the new theory of gravity. Einstein showed how the theory required small adjustments to electrodynamics and thermodynamics. Section 4 of the paper then revealed the concern with the field equation.

Einstein considered a distribution of matter, momentarily at rest, where the gravitational potential c produced by the matter approaches a constant potential at spatial infinity. The different parts of the matter distribution act gravitationally on one another. Gravitational collapse is prevented by attaching the masses to a rigid, massless frame. It follows from the equations of motion (9) that the force density f_i acting on a matter distribution σ momentarily at rest is

$$f_i = -\sigma \frac{\partial c}{\partial x_i} \tag{14}$$

where $i = 1, 2, 3$ so that x_1, x_2, x_3 is x, y, z. The total force acting on the frame at this initial instant is computed by integrating this force density over all space. If we substitute for the matter density σ using the field Eq. (11), we recover

$$\int f_i dV = -\int \sigma \frac{\partial c}{\partial x_i} dV = -\int \frac{\Delta c}{c} \frac{\partial c}{\partial x_i} dV \neq 0 \tag{15}$$

where the integral extends over all of three-space. This integral does not, in general, vanish, Einstein noted. Thus there is a net force acting on the mass-frame system that seeks to set it into motion.

This, Einstein observed, violates the "principle of equality of action and reaction." Alternatively, we might observe that it violates both energy and momentum conservation, since the mass-frame system spontaneously acquires both. Einstein could not disguise his alarm. He wrote (p. 453):

> We have recovered a very questionable result. It is quite enough to arouse doubt over the admissibility of the entire theory developed here. This result certainly indicates a lacuna that lies deeply in the foundation of both our investigations. For it can hardly work out that another equation other than Eq. (10) can be brought into consideration from the expression $(c_0 + ax)$ found for c for a uniformly accelerated system. This [equation] in turn entails Eq. (11) necessarily.

9 Seeking an Escape

Einstein's remarks foreshadow that he will have to give up his pair of field Eqs. (10) and (11). However he was not prepared to take this step without resistance. He sought first to preserve them by modifying other parts of his theory.

The first approach was to consider the fact that the massless frame holding the masses is stressed as it prevents the gravitational collapse of the masses it carries. Earlier work in special relativity had shown that stressed bodies can have unexpected energetic properties. For example, if a stressed body is set in motion, there will be an energy associated with the stress that only appears when the body is in motion.[15] Might there be a gravitational mass associated with the stresses in the frame that somehow preserves the equality of action and reaction? Einstein explored the possibility by considering a mirrored box that contains radiation; and another box containing an ideal gas. In both cases, the walls of the boxes would become stressed in virtue of the pressures exerted on them by the radiation and the gas. However, Einstein concluded, one could not attribute a gravitational mass to the stressed walls. The gravitational mass of the entire system must be determined solely by its total energy. For only then is the equality of inertial and gravitational mass retained. This equality would be violated if an additional gravitational mass were attributed to the stresses in the box walls.[16]

[15] For a survey of these results, see Norton (1992, §9).

[16] Einstein soon returned to the possibility of associating a gravitational mass with stresses in Einstein and Grossmann (1913, §I.7) through the use of the trace T of the stress-energy tensor as the source density in a scalar field equation. Implementing this choice in Nordström's Lorentz covariant gravitation theory led Einstein to a version of the theory that was only conformally flat. It was, as reported in Einstein and Fokker (1914), governed by a field equation $R = \kappa\, T$ where R is the curvature scalar and κ a constant. For further details, see Norton (1992). Giulini (2008) has reconstructed Einstein's argument against Nordström's Lorentz covariant scalar theory of gravity and finds it flawed.

In the second approach, Einstein considered modifying the theory's expressions for the momentum of a moving mass and for the gravitational force by multiplying each by some power in c, the first by c^{α} and the second by c^{β}. Einstein briefly recounts his explorations that showed that these modifications precluded a serviceable dynamics.

10 Modifying the Gravitational Field Equation

Einstein now bowed to the inevitable. The equality of action and reaction could only be preserved, he concluded (p. 455), if his field Eqs. (10) and (11) were modified. We can understand the modification Einstein introduced by reflecting on how ordinary Newtonian gravitation theory and Coulomb electrostatics avoid the problem.

The force density f_i on a charge distribution ρ due to the Coulomb potential φ is given by

$$f_i = -\rho \frac{\partial \varphi}{\partial x_i} \tag{16}$$

The potential is governed by Poisson's equation

$$\Delta\varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\varphi = \sum_{i=1}^{3} \frac{\partial^2 \varphi}{\partial x_i^2} = -k\rho \tag{17}$$

Proceeding as before, we express the force density f_i solely in terms of the potential φ by substituting Poisson's Eq. (17) into (16):

$$\begin{aligned} f_i &= -\rho \frac{\partial \varphi}{\partial x_i} = \frac{1}{k}\left(\sum_{m=1}^{3} \frac{\partial^2 \varphi}{\partial x_m^2}\right)\frac{\partial \varphi}{\partial x_i} \\ &= \frac{1}{k}\sum_{m=1}^{3}\frac{\partial}{\partial x_m}\left(\frac{\partial \varphi}{\partial x_m}\frac{\partial \varphi}{\partial x_i} - \frac{1}{2}\delta_{im}\left(\sum_{n=1}^{3}\frac{\partial \varphi}{\partial x_n}\frac{\partial \varphi}{\partial x_n}\right)\right) = \frac{1}{k}\sum_{m=1}^{3}\frac{\partial t_{im}}{\partial x_m} \end{aligned} \tag{18}$$

What will prove the most important step in this computation is the third equality. It is merely the computation of an identity in φ:

$$\left(\sum_{m=1}^{3} \frac{\partial^2 \varphi}{\partial x_m^2}\right)\frac{\partial \varphi}{\partial x_i} = \sum_{m=1}^{3}\frac{\partial}{\partial x_m}\left(\frac{\partial \varphi}{\partial x_m}\frac{\partial \varphi}{\partial x_i} - \frac{1}{2}\delta_{im}\left(\sum_{n=1}^{3}\frac{\partial \varphi}{\partial x_n}\frac{\partial \varphi}{\partial x_n}\right)\right) \tag{19}$$

The last term in the scope of the divergence operator is the Maxwell stress tensor for the Coulomb field, which is defined as

$$t_{im} = \frac{\partial \varphi}{\partial x_m}\frac{\partial \varphi}{\partial x_i} - \frac{1}{2}\delta_{im}\left(\sum_{n=1}^{3}\frac{\partial \varphi}{\partial x_n}\frac{\partial \varphi}{\partial x_n}\right) \tag{20}$$

Equation (18) shows that the force density f_i equals the divergence of the stress tensor t_{im}. This fact, we can see, preserves the equality of action and reaction in systems of the type Einstein considered. Take a finite system of charges attached to a rigid frame in a field φ whose spatial derivatives $\partial\phi/\partial x_i$ approach zero as we approach spatial infinity. Using a standard computation routinely employed in field theories, Gauss' theorem allows us to compute the ith component of the net force on system of charges F_i through

$$F_i = \int_V \sum_{k=1}^{3}\frac{\partial t_{ik}}{\partial x_k}dv = \int_A \sum_{k=1}^{3} t_{ik}n_k da \tag{21}$$

The first volume integral extends over a volume of space V sufficiently large for it to contain all the charges and such that the first derivatives of the field $\partial\varphi/\partial x_i$ are brought arbitrarily close to zero on its surface A. The second surface integral extends over the surface A only. The quantity n_i is a unit vector normal to the surface. Since the first derivatives of φ can be brought arbitrarily close zero by making V suitably large, the stress tensor t_{ik} can be made arbitrarily small and so also[17] the net force F_i. This force vanishes if we now take the limit as the volume of integration V exhausts all space. Thus the system of charges and rigid frame experiences no net force. The equality of action and reaction is preserved.

Einstein's gravitational field Eq. (11) seems so close in form to the Poisson equation for Newtonian gravity (12) and for Coulomb electrostatics (17) that we can easily imagine that some similar computation is possible that would preserve the equality of action and reaction. Einstein's (11) differs only in the addition of field potential term c in the field equation's source term $kc\sigma$. Yet that additional term is enough to overturn the whole calculation. To see why, we merely need to repeat the electrostatic calculation of (18) in Einstein's gravitation theory. Substituting the field Eq. (11) into the expression (14) for the force on a static mass distribution, we recover:

$$\begin{aligned} f_i &= -\sigma\frac{\partial c}{\partial x_i} = -\frac{1}{kc}\left(\sum_{m=1}^{3}\frac{\partial^2 c}{\partial x_m^2}\right)\frac{\partial c}{\partial x_i} \\ &= \sum_{m=1}^{3}\frac{\partial}{\partial x_m}\left(-\frac{1}{kc}\left(\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_i} - \frac{1}{2}\delta_{im}\left(\sum_{n=1}^{3}\frac{\partial c}{\partial x_n}\frac{\partial c}{\partial x_n}\right)\right)\right) - \frac{1}{2kc^2}\left(\sum_{m=1}^{3}\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_m}\right)\frac{\partial c}{\partial x_i} \end{aligned} \tag{22}$$

[17] A tacit presumption is that t_{ik} approaches zero faster than the area A grows infinite.

where the last equality is an identity. The calculation can *almost* proceed as before. The force density f_i is equal to a divergence, the divergence of a term quadratic in the derivatives of c, and a second term. The quantity within the scope of the divergence operator $\sum_{m=1}^{3} \frac{\partial}{\partial x_m}(\cdot)$ can be provisionally identified as the gravitational field stress tensor: [18]

$$t_{im} = -\frac{1}{kc}\left(\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_i} - \frac{1}{2}\delta_{im}\left(\sum_{n=1}^{3}\frac{\partial c}{\partial x_n}\frac{\partial c}{\partial x_n}\right)\right) \tag{23}$$

We are very close to the goal of writing the force density f_i as a divergence:

$$f_i = -\sigma\frac{\partial c}{\partial x_i} = \sum_{m=1}^{3}\frac{\partial t_{im}}{\partial x_m} \tag{24}$$

However the second superfluous term of the expression in Eq. (22) remains

$$-\frac{1}{2kc^2}\left(\sum_{m=1}^{3}\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_m}\right)\frac{\partial c}{\partial x_i}$$

It precludes us writing the force density as a divergence. Its presence leads to the non-vanishing force Einstein reported in Eq. (15) above.

A short calculation shows that this second, troublesome term can be eliminated if it is absorbed into the gravitational field Eq. (11). This absorption yields the modified gravitational field equation of the second theory of 1912:

$$\Delta c = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)c = k\left(c\sigma + \frac{1}{2kc}\sum_{m=1}^{3}\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_m}\right) \tag{25}$$

This modified field equation solves the dynamical problem. Using it, the force density f_i can be written as the divergence of a tensor, t_{im} of Eq. (23). An argument analogous to that of the electrostatic case shows the equality of action and reaction is preserved. A bonus, possibly unexpected, is that Einstein could show that the additional term in the modified field equation

$$\frac{1}{2kc}\sum_{m=1}^{3}\frac{\partial c}{\partial x_m}\frac{\partial c}{\partial x_m}$$

[18] Einstein's expression for this stress tensor in (1912a, p. 456) omits the leading minus sign. I believe this is a typographical error in Einstein's paper.

is equal to the energy density of the gravitational field. Einstein now had the appealing result that ordinary matter density σ and the gravitational field energy contribute equally, in arithmetic summation, as the source of the gravitational field:

$$\Delta c = k\,(\text{ordinary matter density} + \text{energy density of the gravitational field}) \tag{26}$$

With this modification, Einstein's theorizing now moved to non-linear field equations, which would be an enduring feature of his development of general relativity and his subsequent unified field theory.

Einstein identified the peril to his theory quite rapidly, sometime between the writing of the first paper (Einstein 1912a) and the second (Einstein 1912b). How did he find it so quickly? Its presence is obvious once one tries the calculation of Eq. (22) above. Why would Einstein try such a calculation? A striking juxtaposition may answer this last question. The first two sections of Einstein (1912b) tackle a mundane exercise required by the new theory. Einstein asks how electrodynamics must be modified to remain compatible with the new theory of gravity. Einstein shows that all that is required is the addition of factors of c in several places. He then proceeds to check that the resulting modified theory retains the conservation of energy and the conservation of momentum. The first computation involves recovering an expression for the electromagnetic field energy density. The second computation leads Einstein to write a modified expression for the Maxwell stress tensor and to show that the modified expression allows retention of the conservation of momentum. The corresponding computation for the gravitational stress tensor of his new theory is the failure that Einstein proceeds to report and that leads to the need for a modification of his gravitational field equation from Eqs. (11) to (25).

11 Conflicting Heuristics

This modification was not just an ad hoc expedient. As we shall see shortly, it embodies a procedure that Einstein could and would use again. It proved to be an invaluable heuristic. The difficulty for Einstein, however, was that this heuristic contradicted the primary heuristic that had played a dominant role in Einstein's thinking on gravity since 1907: the principle of equivalence.

To see the conflict, take what Einstein (1912b, p. 456) correctly gave as an equivalent form of the modified field Eq. (25):

$$\Delta\left(\sqrt{c}\right) = \frac{k}{2}\sqrt{c}\,\sigma \tag{27}$$

For the source free ($\sigma = 0$) case of a unidirectional field such as might be produced by unidirectional acceleration, in analogy with Eq. (6), we recover

$$\sqrt{c} = \sqrt{c_0} + ax \tag{28}$$

The difficulty is that uniform acceleration in special relativity produces (6) and not (28). That is, the gravitational field of the principle of equivalence, produced by uniform acceleration, is not a gravitational field admitted by the modified field Eq. (25)/(27).

One might wonder if there is some scope for modifying the transformation Eq. (1) used to produce the field represented by (6). Using later ideas, we can see that this is not possible, unless Einstein is prepared to make much more sweeping changes to this theory. If we assume that the spacetime geometry is given by the line element (8), then the function $c(x, y, z)$ in Eq. (8) can vary at most linearly with the spatial coordinates x, y, and z. This linearity is shown in Appendix 1 by the analysis leading to Eq. (42).

Einstein (1912b, pp. 455–56) reported his reluctance to adopt the modified field equation:

> Therefore I decided with difficulty to take this step, since with it the foundation of the unconditional principle of equivalence is lost. It appears that the latter can only be retained for infinitely small fields.

Presumably the restriction is to infinitely small intervals of space in the direction of the x coordinate. For then the non-linear dependency of c on x in Eq. (28) can be approximated in the infinitely small interval by the linear dependency of Eq. (6).

All was not lost entirely, Einstein continued. For his derivation of the equations of motion (9) and the modification to the equations of electrodynamics from the principle required only that his transformation Eq. (2) can be applied to infinitely small spaces. He suggested that the transformation Eq. (2) be replaced by the more general equations:

$$\tau = ct \quad \xi = x + \frac{1}{2} c \frac{\partial c}{\partial x} t^2 \quad \eta = y \quad \zeta = z \tag{29}$$

where c is an arbitrary function of x.

While the outcome was clearly painful for Einstein, there is an unmistakable conclusion concerning Einstein's heuristics. Einstein's first tier and most visible heuristic of the principle of equivalence conflicted with the less visible, second tier heuristic of momentum conservation. The second tier heuristic wins. It is, in the end, a more powerful guide that cannot be overruled.

12 A Method Reused: The Derivation of the "Entwurf" Field Equations

The procedure Einstein used in 1912 to correct his gravitational field equation was not something merely to be used once. It could be reused in different contexts. That is, for Einstein, it was a method. We know this because he goes to some pains to tell us. The word "method" is his, as we shall see below.

In their "Entwurf" paper of 1913, Einstein and his collaborator, his mathematician friend Marcel Grossmann, published an almost complete version of the general theory of relativity (Einstein and Grossmann 1913). What was missing were the now celebrated Einstein gravitational field equations.[19] In their place, Einstein offered field equations of limited covariance. In his physical part of their joint paper, Einstein addressed the problem of identifying these equations. Following a now familiar approach, he posited that these gravitational field equations would have the form

$$\Gamma^{\mu\nu} = kT^{\mu\nu} \tag{30}$$

where $T^{\mu\nu}$ is the stress-energy tensor for ordinary matter and k is a constant. The gravitation tensor, $\Gamma^{\mu\nu}$, is a quantity constructed from the metric tensor $g_{\mu\nu}$ and its first and second coordinate derivatives. Unlike his later theory, this tensor was permitted only limited covariance. In the case of a spacetime whose metric differed only in small quantities from that of a Minkowski spacetime, Einstein specified (pp. 13–15) that the gravitation tensor would have the form (in more modern notation):

$$\Gamma^{\mu\nu} = \frac{\partial}{\partial x^{\alpha}} \left(g^{\alpha\beta} \frac{\partial g^{\mu\nu}}{\partial x^{\beta}} \right) + \begin{array}{c}\text{further terms that vanish in the}\\ \text{formation of the first approximation}\end{array} \tag{31}$$

How could these further terms be found?

Einstein saw that his situation was quite similar to that of 1912. One could conceive his first gravitational field Eq. (11) merely as an approximation to the correct equation, merely lacking the higher order terms introduced in the second gravitational field Eq. (25). Einstein had found these higher order terms by requiring that substitution of the force density Eq. (16) into the field equation must produce an identity from which the conservation of momentum could be recovered. Without mentioning the embarrassing retraction of 1912, Einstein now sought to employ the same method in his new "Entwurf" theory. He was concerned to convey clearly to the reader the method that would be used. To do so, he recapitulated the analysis given above in Sect. 10 for the familiar case of electrostatics.[20] I quote him at length (p. 14):

> The momentum energy law will serve us in the discovery of these terms. So that the method used is clearly delineated, I now want to apply it to a generally known example.

[19] The story of their rejection of generally covariant field equations has been told in abundance elsewhere. See Stachel (1989), Norton (1984), and Renn (2007).

[20] A curious omission is that Einstein never states the key point explicitly: that conservation of momentum is assured by the existence of the Maxwell stress tensor. Perhaps he assumed it would be obvious to the reader?

In electrostatics, $-\frac{\partial\varphi}{\partial x_\nu}\rho$ is the νth component of the momentum per unit volume imparted to matter, in case φ signifies the electrostatic potential, ρ the electric [charge] density. A differential equation is sought for φ of such a kind that the momentum law is always satisfied. It is well-known that the equation

$$\sum_\nu \frac{\partial^2\varphi}{\partial x_\nu^2} = \rho$$

solves the exercise. That the momentum law is satisfied follows from the identity

$$\sum_\mu \frac{\partial}{\partial x_\mu}\left(\frac{\partial\varphi}{\partial x_\nu}\frac{\partial\varphi}{\partial x_\mu}\right) - \frac{\partial}{\partial x_\nu}\left(\frac{1}{2}\sum_\mu\left(\frac{\partial\varphi}{\partial x_\mu}\right)^2\right) = \frac{\partial\varphi}{\partial x_\nu}\sum_\mu\frac{\partial^2\varphi}{\partial x_\mu^2}\left(= -\frac{\partial\varphi}{\partial x_\nu}\cdot\rho\right)$$

Therefore if the momentum law is satisfied, for each ν an identical equation of the following construction must exist: on the right hand side is $-\frac{\partial\varphi}{\partial x_\nu}$ multiplied by the left hand side of the differential equation. On the left hand side of the identity is a sum of differential quotients.

If the differential equation for φ were not yet known, then the problem of its discovery may be reduced to that of the discovery of this identical equation. What is essential for us now is the knowledge that this identity may be derived *if one of the terms appearing in it is known*. [Einstein's emphasis] One has nothing more to do than to apply repeatedly the rule for differentiation of a product in the form

$$\frac{\partial}{\partial x_\nu}(uv) = \frac{\partial u}{\partial x_\nu}v + \frac{\partial v}{\partial x_\nu}u$$

and

$$u\frac{\partial v}{\partial x_\nu} = \frac{\partial}{\partial x_\nu}(uv) - \frac{\partial u}{\partial x_\nu}v$$

and finally to place terms that are differential quotients on the left hand side and the remaining [terms] on the right hand side. If one proceeds, f[or] e[xample] from the first term of the above identity, one obtains the sequence

$$\sum_\mu \frac{\partial}{\partial x_\mu}\left(\frac{\partial\varphi}{\partial x_\nu}\frac{\partial\varphi}{\partial x_\mu}\right) = \sum_\mu\left(\frac{\partial\varphi}{\partial x_\nu}\cdot\frac{\partial^2\varphi}{\partial x_\mu^2}\right) + \sum_\mu\left(\frac{\partial\varphi}{\partial x_\nu}\cdot\frac{\partial^2\varphi}{\partial x_\nu\,\partial x_\mu}\right)$$

$$= \frac{\partial\varphi}{\partial x_\nu}\cdot\sum_\mu\frac{\partial^2\varphi}{\partial x_\mu^2} + \frac{\partial}{\partial x_\nu}\left\{\frac{1}{2}\sum_\mu\left(\frac{\partial\varphi}{\partial x_\mu}\right)^2\right\}$$

from which the above identity follows through rearrangement.

Einstein now proceeded to use this method to derive the gravitational field equations of his "Entwurf" theory. The derivation was essentially just the derivation of the second gravitational field equation of 1912, but now promoted to the more complicated context of the "Entwurf" theory. In place of the single gravitational potential c was the multi-component metric tensor $g_{\mu\nu}$. In place of momentum conservation and the Maxwell stress tensor was the requirement of conservation of energy-momentum and the stress-energy tensor of the gravitational field.

The resulting gravitation tensor $\Gamma^{\mu\nu}$ is given in Appendix 2 below as Eq. (43). The promoted computations are considerably more complicated than those of the 1912 theory. Grossmann's (1913, pp. 37–38) part contains the derivation of the essential identity, which covers two journal pages. The details of these formulae are unilluminating for our present interests and I will spare the reader parading them.

13 Conflicting Heuristics Again

While the promotion of the method of 1912 had now provided Einstein with a unique set of gravitational field equations for his new "Entwurf" theory, the conflict of heuristics present in 1912 remained and in a more damaging form. The principle of equivalence had assured Einstein that uniform acceleration produces a homogeneous gravitational field. We saw above that Einstein's modified field equation of 1912 no longer identified this acceleration field in its totality as a gravitational field. The best Einstein could say was that infinitely small parts of the field were identified individually as a gravitational field. In the "Entwurf" theory, this last slender thread to the principle was broken. For the only static spacetimes with the line element (8) allowed by the source free "Entwurf" gravitational field equations were those with $c =$ constant. (See Appendix 2 Eq. (46).) That is merely the spacetime of special relativity, Minkowski spacetime.

The immediate problem was that Einstein could not present the "Entwurf" theory as realizing the idea implicit in the principle of equivalence. For in this new theory, uniform acceleration did not produce a gravitational field that was recognized by the theory's gravitational field equations. Hence, as noted already in Norton (1985, §4.3), during the time of the "Entwurf" theory, Einstein tended to avoid detailed discussion of the principle of equivalence.

In his part of Einstein and Grossmann (1913), the principle ("Äquivalenz-Hypothese") is introduced (p. 3) with the restriction to homogeneous gravitational fields of infinitely small extension. It is recalled subsequently (§1) only as the basis of the 1912 theory, which is summarized briefly. In his later Einstein (1913, pp. 1254–1255), the principle is presented as a vividly developed thought experiment concerning physicists who awaken from a drugged sleep in a closed, accelerating box. Einstein does not, however, develop the specific results such as the line element (4) above. Soon after, Einstein and Grossmann (1914, p. 216) reaffirm

> The whole theory proceeds from the conviction that all physical processes in a gravitational field play out in exactly the same way as the corresponding processes play out without a gravitational field, in case one relates them to an appropriately accelerated (three dimensional) coordinate system. ("Äquivalenzhypothese")

It is notable that Einstein and Grossman leave open just what form the "appropriate" acceleration can take. They fail to specify the uniform acceleration and homogeneous gravitational fields of Einstein's earlier formulations and those of his writings after 1915. In November 1914, Einstein (1914) published a definitive review article on the latest form of his theory. The principle of equivalence is now absent in name from the introductory discussion. Instead, Einstein reflects on rotational motion and urges (p. 1032) that the centrifugal field appearing in a rotating frame of reference should be conceived as a gravitational field.[21]

After November 1915, when Einstein had finally secured a generally covariant theory, he could once again conceive of the field of uniform acceleration as gravitational. The principle of equivalence was restored in its original form to its original prominence in Einstein's accounts of his theory. It appears in the introductory discussion (§2) of his new review article (Einstein 1916); and in Ch.XX of Einstein's (1920b) popular book on relativity, whose preface is dated December 1916.

The hiatus in discussion of the principle of equivalence coincides with the time of the "Entwurf" theory. Thus is it natural to suppose that Einstein knew that the original principle failed completely in his theory. Unfortunately Einstein never explicitly acknowledged the failure. What complicates the problem is that some of Einstein's narratives (cited above) still include it. What deepens the problem is that Einstein repeatedly employed a spacetime with a line element (8) to represent the gravitational field outside a spherically symmetric body, such as was assumed for the sun. The difficulty is that this field must conform with the source free gravitation field equations and, as shown in Appendix 2, these equations admit nothing but a flat Minkowski spacetime for a spacetime with this line element.

While Einstein's silence makes it impossible for us to be certain, I think it most plausible that Einstein knew of the problem but found it expedient to remain silent about it. For once a successful theory has been achieved, what could be gained by announcing incompatibilities between the theory and the specifics of the ideas that led to it? If uniform acceleration does not produce a gravitational field in the theory, then other accelerations might; and Einstein mistakenly believed this to be the case for rotation. As to the applicability of the line element (8) to the spacetime surrounding the sun, it is notable that Einstein's derivations all employ approximations.[22] Thus the negative result of Appendix 2 below could be avoided if the line element of these spacetimes had the form (8) only approximately, that is, to the order of the approximation of his calculations. I find it most plausible that this was Einstein's view.

There is evidence that Einstein knew that static gravitational fields, such as that of the sun, admitted deviations from spatial flatness that were non-zero in the

[21] At this time, as Janssen (1999) recounts in some detail, through a calculation error, Einstein had convinced himself that this centrifugal field is a solution of the "Entwurf" gravitational field equations.

[22] See for example Einstein (1913, §8) and Einstein (1914, §17). In a letter of March 19, 1915 (Schulmann et al. 1998, Doc. 63.), Einstein sought to reassure Erwin Freundlich that the spacetime surrounding the sun has the metric associated with (8)/(36). Einstein presented a short proof that demonstrates the result only in low order approximation.

second order of smallness. The most direct evidence comes in a draft manuscript of calculations co-authored by Einstein and his friend Michele Besso, mostly in mid-1913 (Klein et al., 1995, Doc. 14). They compute the gravitational field of the sun to second order in the "Entwurf " theory and, on a page in Einstein's hand, non-zero second order deviations are recovered.[23]

What is regrettable is that Einstein does not directly affirm these deviations in his publications from the time. Einstein and Grossmann (1913, p. 7) present the line element (8) (in the equivalent form of the tensor (36)) as applying to static gravitational fields "of the previously considered type." This presumably refers to those of the earlier 1912 theory. If Einstein intended the remark not to apply to the present "Entwurf" theory as well, only the most perspicacious of readers could have realized it.

In November 1915, after Einstein had returned to generally covariant gravitational field equations, the error was discovered in the context of Einstein's successful explanation of the anomalous motion of Mercury. He then remarked (Einstein 1915b, p. 834) on the surprising[24] appearance of non-constant components like g_{11}, g_{22}, and g_{33} in the metric field of the sun: "the [non-constancies of the] components g_{11} to g_{33} differ from zero already *in magnitudes of the first order.* [my emphasis]" This emphasized phrase might not be needed, unless Einstein already had expected such deviations only to be of higher order.

14 The Method Lives On

After the "Entwurf" paper, the essential ideas behind the method of generating field equations did not disappear, but merely receded. They were absorbed into Einstein's analyses and, while no longer explicitly delineated, continued to exercise a controlling influence on his theorizing.

The gravitational field equations of the "Entwurf" theory were not generally covariant. The pressing problem for Einstein in 1913 and 1914 was to determine the extent of his new theory's covariance. The ideas behind the method of 1912 and 1913 now became the vehicle for determining this extent. To this end, Einstein and Grossmann (1914, p. 217) wrote the "Entwurf" field equations (in modernized notation) as:

$$\frac{\partial}{\partial x^{\alpha}}\left(\sqrt{-g}g^{\alpha\beta}g_{\sigma\mu}\frac{\partial g^{\mu\nu}}{\partial x^{\beta}}\right)=\kappa\sqrt{-g}\left(T_{\sigma}^{\nu}+t_{\sigma}^{\nu}\right) \tag{32}$$

[23] See Equations 40 and 42 in Klein et al. (1995, p. 370) and the associated editorial discussion on p. 349.

[24] The word surprise is Einstein's from a letter to Michele Besso of December 10, 1915: "You will be surprised by the appearance of the g_{11} ... g_{33}." (Schulmann et al. 1998, Doc. 162, p. 218).

where $t_\sigma{}^\nu$ is an expression quadratic in first derivatives of the metric tensor and identified as the stress-energy tensor of the gravitational field. The conservation of energy and momentum was written as

$$\frac{\partial}{\partial x^\nu}\left(\sqrt{-g}\left(T_\sigma^\nu + t_\sigma^\nu\right)\right) = 0 \tag{33}$$

Following the earlier method, we should expect that substituting (32) into (33) yields an identity in the metric tensor $g_{\mu\nu}$. The resulting identity in $g_{\mu\nu}$ is

$$B_\sigma = \frac{\partial^2}{\partial x^\nu \partial x^\alpha}\left(\sqrt{-g}g^{\alpha\beta}g_{\sigma\mu}\frac{\partial g^{\mu\nu}}{\partial x^\beta}\right) = 0 \tag{34}$$

This identity took on a new significance. It could only be expected to hold in coordinate systems in which the original Eqs. (32) and (33) held. Einstein and Grossmann could now use the identity as the condition that picks out just those coordinate systems in which the "Entwurf" theory held.

This "adapted coordinate condition," as they called it, became a central feature of the development of the "Entwurf" theory. Einstein and Grossmann (1914) and Einstein (1914) developed a variational formalism for the "Entwurf" theory. A major goal of the formalism was to demonstrate that this adapted coordinate condition did characterize precisely the extent of covariance of the theory and that it was the maximum covariance permitted by Einstein's original interpretation of the hole argument.

When Einstein returned to general covariance and formulated the now familiar generally covariant gravitational equations, the ideas behind this repurposed method and the variational formalism persisted. The major difference was that the identity replacing (34) no longer picked out just those few coordinate systems in which the theory held. For under general covariance, the final theory held in all coordinate systems. Thus the replacement identity must hold in all coordinate systems. It was recognized later to be none other that the contracted Bianchi identity:

$$\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu} = 0 \tag{35}$$

where $R^{\mu\nu}$ is the Ricci tensor.

Einstein's original method of 1912 and 1913 now survives as the most familiar means of arriving at the gravitational field equations of general relativity. It proceeds by arguing, as did Einstein (1923, pp. 92–93), that the gravitational field equations have the form

$$G^{\mu\nu} = kT^{\mu\nu}$$

The generally covariant gravitation tensor $G^{\mu\nu}$ is formed from the metric tensor and its first and second derivatives; and it is linear in the second derivatives. It follows that $G^{\mu\nu}$ must be a linear combination of $R^{\mu\nu}$ and $g^{\mu\nu}$R. If conservation of energy momentum

$$T^{\mu\nu}{}_{;\nu} = 0$$

is to be maintained, this linear combination must have a vanishing covariant divergence. We read from (35) that the gravitation tensor is what is now called the Einstein tensor:

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R$$

15 Einstein's Two Tier Heuristics

To return to our starting point, how are we to think of the heuristics that guided Einstein to his general theory of relativity? His starting point in 1907 was the principle of equivalence. There can be no doubt that Einstein held firmly to the idea that this principle was the foundation from which he proceeded, even as the principle delivered results in contradiction with his evolving theory. Here is how Einstein recalled the situation in a letter of September 12, 1950, to Laue. At issue was the fact that the Riemann curvature tensor vanishes in the rotating coordinate system adapted to a rotating disk in Minkowski spacetime. Einstein replied[25]

> It is true that in that case the R_{iklm}vanish, so that one could say: "there is no gravitational field present." However, what characterizes the existence of a gravitational field from the empirical standpoint is the non-vanishing of the $\Gamma_{ik}{}^{l}$ [coefficents of the affine connection], not the non-vanishing of the R_{iklm}. If one does not think intuitively in such a way, one cannot grasp why something like curvature should have anything at all to do with gravitation. In any case, no reasonable person would have hit upon such a thing. The key for understanding of the equality of inertial and gravitational mass is missing.

In retrospect, we can see the most important idea that the principle of equivalence delivered to Einstein. It was, as is argued in Norton (1985, §12), that the Minkowski spacetime of special relativity was not to be conceived as a gravitation free spacetime. Rather gravitation was already present in it as a special case. That gave Einstein the crucial clue that a gravitation theory could be constructed, not by adding a gravitational field to that spacetime, but generalizing the structures already present in Minkowski spacetime.

The difficulty was that the principle of equivalence gave Einstein more than this vital clue. It also delivered gravitational fields to Einstein that contradicted his evolving theories of 1912 and 1913. If the principle of equivalence was inviolable, Einstein would have had to abandon these theories. He did not; the principle proved

[25] As quoted in Norton (1985, §11).

dispensable. Rather he first reduced the principle of equivalence in 1912 to a weak version that obtained only in the infinitely small and then in 1913 and 1914 to a vaguer guide with an imprecisely circumscribed expression. The principle may have taken pride of place in his overarching conceptions, but it enjoyed no such prominence in the practicalities of his theorizing.

Instead Einstein could proceed with quite definite theories because a second tier of heuristics were still guiding him. In the account above, one has been singled out as having special importance.[26] It is the idea that the gravitational field equations must conform with energy and momentum conservation. Unlike the principle of equivalence, that demand was inviolable. It provided a method that guided Einstein to quite specific field equations in 1912 and in 1913 and persists in modern presentations of general relativity.

Is this example of a two-tiered structure of heuristics in Einstein's work exceptional? A second, related example has been explored in some detail by Janssen and Renn (2006). In November 1915, Einstein (1915a) reported to the Prussian Academy that he had abandoned his "Entwurf" theory. He presented in its place a new theory of near general covariance that would shortly be extended to full general covariance. Einstein made clear that, once he had lost faith in his earlier theory, considerations of covariance were his primary guide: (p. 778)

> Thus I came back[27] to the demand of a more general covariance of the field equations, from which I had departed three years ago, when I worked together with my friend Grossmann, only with a heavy heart.

His reflections devolved into a poetic tribute to the mathematical methods associated with general covariance (p. 779)

> Hardly anyone who has truly understood it can resist the charm of this theory; it signifies a real triumph of the method of the general differential calculus, founded by Gauss, Riemann, Christoffel, Ricci and Levi-Civita.

Janssen and Renn, however, have pointed out that the theory then presented by Einstein could be produced by making a small adjustment to the variational formulation of the "Entwurf" theory. A derivative of the metric tensor would be replaced by a Christoffel symbol, otherwise leaving the formalism unchanged. The cogency of the ensuing theory was assured by the results of the earlier formulation. In particular, the modified theory would be assured to conform with the conservation of energy and momentum.

There is no reason to doubt that Einstein conveyed accurately his perception of the overriding importance of covariance considerations. That would be a natural way for him to recall his recognition that the modified theory was the same as one of near general covariance, recoverable from the Riemann tensor. However it obscures how powerfully his further demands constrained his choices.

[26] Another essential requirement was that his new gravitation theory revert to Newtonian gravitation theory in the case of weak, static gravitational fields.

[27] "So gelangte ich ... zurück ..."

We see in this example a similar double tiered structure of heuristics. Covariance considerations loomed large in Einstein's thinking as the first tier. However they were quite dispensable. Einstein had abandoned them in 1913 with the formulation of the "Entwurf" theory, whose covariance properties were then unclear. Considerations such as the energy momentum conservation and the Newtonian limit, however, were inviolable and formed the second tier that continued to guide and circumscribe his theorizing. In November 1915, Einstein could return to more general covariance precisely because he had in hand a formalism that preserved the demands of this second tier.

16 Conclusion

It is tempting to say that Einstein did not really need the principle of equivalence to guide him to general relativity. The crucial clue that Minkowski spacetime is already gravitational could have been gleaned from a widely known fact, itself brought to prominence by Einstein's work. It is the remarkable equality of inertial and gravitational mass in Newtonian theory. This equality leads to the result that trajectories of bodies in free fall are independent of their mass. They are, in retrospect, tracing out for us the geometry of a curved spacetime associated with gravity. Might that have been enough to guide Einstein or another theorist to general relativity?

Of course, when our concern is the discovery of a theory as exceptional in relation to what went before as general relativity, it is foolhardy to try to imagine how things could have been otherwise. I will not persist. We saw above that Einstein insisted that without the principle of equivalence "no reasonable person" could have found general relativity. However, just as I cannot really know how it would have been if things were otherwise, is not the same true for Einstein?

Acknowledgment I thank Michel Janssen for helpful discussion in general and especially relating to Section 13; and Dennis Lehmkuhl and Tilman Sauer for comments and corrections on an earlier draft.

Appendix 1: Computing Spacetime Curvature in Einstein's 1912 Theory

Einstein's 1912 theory of static gravitational fields attributed properties to space and time equivalent to spacetimes of his later general theory of relativity with a spacetime metric:

$$g_{\mu\nu} = \begin{bmatrix} -1 & & & \\ & -1 & & \\ & & -1 & \\ & & & c^2(x, y, z) \end{bmatrix} \quad g^{\mu\nu} = \begin{bmatrix} -1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 1/c^2(x, y, z) \end{bmatrix} \tag{36}$$

where the spacetime coordinates are $(x, y, z, t) = (x^1, x^2, x^3, x^4)$ and Greek indices μ, ν take values 1, 2, 3, and 4. In (36), c is a function of x, y, and z, but not t. Following the notational conventions of Einstein (1923, p. 79), we write the coefficients of the connection as

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2} g^{\sigma\alpha} \left(\frac{\partial g_{\mu\alpha}}{\partial x^{\nu}} + \frac{\partial g_{\nu\alpha}}{\partial x^{\mu}} - \frac{\partial g_{\mu\nu}}{\partial x^{\alpha}} \right) \tag{37}$$

where summation over repeated indices is implied. Substituting Eq. (36) in Eq. (37), the only non-zero terms are

$$\Gamma^{i}_{44} = c \frac{\partial c}{\partial x^{i}} \quad \Gamma^{4}_{i4} = \Gamma^{4}_{4i} = \frac{1}{c} \frac{\partial c}{\partial x^{i}} \tag{38}$$

where a Latin index $i = 1, 2, 3$, is used to identify the spatial coordinates $(x, y, z) = (x^1, x^2, x^3)$. The Ricci tensor, as given by Einstein (1923, p. 85), is

$$R_{\mu\nu} = -\frac{\partial \Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} + \frac{\partial \Gamma^{\alpha}_{\mu\alpha}}{\partial x^{\nu}} + \Gamma^{\alpha}_{\mu\beta} \Gamma^{\beta}_{\nu\alpha} - \Gamma^{\alpha}_{\mu\nu} \Gamma^{\beta}_{\alpha\beta} \tag{39}$$

Using the values of (38), after some calculations, the Ricci tensor reduces to[28]

$$\begin{aligned} R_{44} &= -c\Delta c = -c \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) c \\ R_{ik} &= \frac{1}{c} \frac{\partial^2 c}{\partial x^i \partial x^k}, \text{ so that } R_{11} = \frac{1}{c} \frac{\partial^2 c}{\partial x^2}, R_{12} = \frac{1}{c} \frac{\partial^2 c}{\partial x \partial y}, \text{ etc.} \\ R_{i4} &= R_{4i} = 0 \end{aligned} \tag{40}$$

Finally, the Riemann curvature scalar is

$$R = g^{\mu\nu} R_{\mu\nu} = \frac{1}{c^2} R_{44} - R_{11} - R_{22} - R_{33} = -\frac{2}{c} \Delta c \tag{41}$$

[28]These formulae are accurate to all orders. They differ from Stachel's (1989, p. 67) formulae, which are computed only, in Stachel's expression, in "linearized approximation."

Einstein's source free gravitational field equations of 1915, $R_{\mu\nu} = 0$, lead to highly restricted results. The spatial part, $R_{ik} = 0$, alone is sufficient to ensure that c depends at most linearly on the spatial coordinates x, y, and z. That is

$$c\,(x, y, z) = A + Bx + Cy + Dz \tag{42}$$

where A, B, C, and D are constants. Equation (42) also applies to the special case of flat spacetime, when the Riemann curvature tensor vanishes. For in that case, its first contraction must also vanish, $R_{\mu\nu} = 0$.

Had Einstein set his source density in his field Eq. (11), $\Delta c = kc\sigma$, equal to the trace of the stress-energy tensor of ordinary matter, that is, $\sigma = T$, then it follows from (41) that the field Eq. (11) would be equivalent to

$$-R = (2/c)\,\Delta c = 2k\sigma = 2kT$$

Appendix 2: Computing the Gravitation Tensor of the Einstein–Grossmann Theory for a Static Gravitational Field

The gravitation tensor of limited covariance of Einstein and Grossmann (1913, p. 15) is given in more modern notation as:

$$\begin{aligned}\Gamma^{\mu\nu} = {} & \tfrac{1}{\sqrt{-g}}\tfrac{\partial}{\partial x^\alpha}\left(\sqrt{-g}g^{\alpha\beta}\tfrac{\partial g^{\mu\nu}}{\partial x^\beta}\right) - g^{\alpha\beta}g_{\tau\sigma}\tfrac{\partial g^{\mu\tau}}{\partial x^\alpha}\tfrac{\partial g^{\nu\rho}}{\partial x^\beta} \\ & + \tfrac{1}{2}g^{\alpha\mu}g^{\beta\nu}\tfrac{\partial g_{\tau\rho}}{\partial x^\alpha}\tfrac{\partial g^{\tau\rho}}{\partial x^\beta} - \tfrac{1}{4}g^{\mu\nu}g^{\alpha\beta}\tfrac{\partial g_{\tau\rho}}{\partial x^\alpha}\tfrac{\partial g^{\tau\rho}}{\partial x^\beta}\end{aligned} \tag{43}$$

Evaluating this tensor for the static spacetimes (36), as conceived in the Einstein and Grossmann theory, we find the only non-zero derivatives of the metric tensor are:

$$\frac{\partial g_{44}}{\partial x^i} = 2c\frac{\partial c}{\partial x^i} \qquad \frac{\partial g^{44}}{\partial x^i} = -\frac{2}{c^3}\frac{\partial c}{\partial x^i} \tag{44}$$

where $i = 1, 2, 3$. After some straightforward computations, we recover

$$\Gamma^{44} = \frac{2}{c^3}\sum_{i=1}^{3}\frac{\partial^2 c}{\partial\left(x^i\right)^2} - \frac{1}{c^4}\sum_{i=1}^{3}\left(\frac{\partial c}{\partial x^i}\right)^2 \quad \Gamma^{ii} = -\frac{2}{c^2}\left(\frac{\partial c}{\partial x^i}\right)^2 + \frac{1}{c^2}\sum_{i=1}^{3}\left(\frac{\partial c}{\partial x^i}\right)^2$$

$$\Gamma^{ik}{}_{(i\neq k)} = -\frac{2}{c^2}\frac{\partial c}{\partial x^i}\frac{\partial c}{\partial x^k} \qquad \Gamma^{i4} = \Gamma^{4i} = 0 \tag{45}$$

where all summations are explicit. No summations are implied. For the source free case, the gravitational field equations of the theory are $\Gamma^{\mu\nu} = 0$. The component equation $\Gamma^{44} = 0$ corresponds to the source free form of the second field Eq. (25) of Einstein's second theory of 1912. The remaining component equations, however, have terms in the first derivatives $\partial c/\partial \mathrm{x}^i$ only. The three equations $\Gamma^{ii} = 0$, for $i =$ 1, 2, 3, are sufficient to force $\partial c/\partial \mathrm{x}^i = 0$ for $i = 1, 2, 3$. That is, we must have

$$c\,(x, y, z) = \text{constant} \tag{46}$$

References

Einstein, Albert (1907) "Das Relativitätsprinzip und die aus demselben gezogenen Folgerungen," *Jahrbuch der Radioaktivität und Electronik*, **4**, pp. 411–462.

Einstein, Albert (1911) "Über den Einfluss des Schwerkraft auf die Ausbreitung des Lichtes," *Annalen der Physik*, **35**, pp. 898-908.

Einstein, Albert (1912a) "Lichtgeschwindigkeit und Statik des Gravitationsfeldes," *Annalen der Physik*, **38**, pp. 355-69.

Einstein, Albert (1912b) "Zur Theorie des statischen Gravitationsfeldes," *Annalen der Physik*, **38**, pp. 443-58.

Einstein, Albert (1913) "Zum gegenwärtigen Stande des Gravitationsproblems," *Physikalische Zeitschrift*, **14**, pp. 1249-1262.

Einstein, Albert (1914) "Die formale Grundlage der allgemeinen Relativiätstheorie," *Königlich Preussische Akademie des Wissenschaften* (Berlin), 1914, pp. 1013-1085.

Einstein, Albert (1915a) "Zur allgemeinen Relativitätstheorie," *Königlich Preussische Akademie des Wissenschaften* (Berlin), 1915, pp. 778-85.

Einstein, Albert (1915b) "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie," *Königlich Preussische Akademie des Wissenschaften* (Berlin), 1915, pp. 831-39.

Einstein, Albert (1916) "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik*, **49**, pp. 769-822.

Einstein, Albert (1920a) "Grundgedanken und Methoden der Relativitätstheorie, in ihrer Entwicklung dargestellt," Doc. 31 in, M. Janssen et al. eds, *The Collected Papers of Albert Einstein. Volume 7*. Princeton: Princeton University Press, 2002.

Einstein, Albert (1920b) *Relativity: The Special and the General Theory*. 3rd ed. London: Methuen.

Einstein, Albert (1923) *The Meaning of Relativity.* Princeton: Princeton University Press.

Einstein, Albert (1933) "Notes on the Origin of the General Theory of Relativity." In *Ideas and Opinions*, pp. 285–290. Translated by Sonja Bargmann. New York: Crown, 1954.

Einstein, Albert and Fokker, Adriaan D. (1914) Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls, Annalen der Physik, **44**, pp. 321–28.

Einstein, Albert and Grossmann, Marcel (1913) *Entwurf einer Verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig u. Berlin: B. G. Teubner.

Einstein, Albert and Grossmann, Marcel (1914) "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie," *Zeitschift für Mathematik und Physik,**63**, pp. 215-25.*

Einstein, Albert and Rosen, Nathan (1935) "The Particle Problem in the General Theory of Relativity," *Physical Review*, **48**, pp. 73-77.

Giulini, Domenico (2008) "What is (not) wrong with scalar gravity?" *Studies in the History and Philosophy of Modern Physics*, **39**, pp. 154–180.

Janssen, Michel (1999) "Rotation as the Nemesis of Einstein's Entwurf Theory," pp. 127-157 in Hubert Goenner et al. (eds.), *The Expanding Worlds of General Relativity. Einstein Studies.* Vol. 7 Boston: Birkhäuser.

Janssen, Michel (2014) "'No Success Like Failure ...': Einstein's Quest for General Relativity, 1907–1920." pp. 167–227 in Michel Janssen and Christoph Lehner, eds. *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press,.

Janssen, Michel and Renn, Jürgen (2007) "Untying the Knot: How Einstein Found his Way Back to the Field Equations Discarded in the Zürich Notebook," pp. 839-925 in Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 2 *Einstein's Zurich Notebook: Commentary and Essays.* Dordrecht: Springer.

Klein, Martin J.; Kox, A. J.; Renn, Jürgen; and Schulman, Robert, eds., (1995) *Collected Papers of Albert Einstein. Volume 4. The Swiss Years: Writings 1912-1914.* Princeton: Princeton University Press.

Lorentz, Hendrik A. (1920) *The Einstein Theory of Relativity: A Concise Statement*. New York: Brentano's publishers.

Norton, John D. (1984) "How Einstein Found His Field Equations: 1912-1915," *Historical Studies in the Physical Sciences*, **14**, pp. 253-315.

Norton, John D. (1985) "What was Einstein's Principle of Equivalence?" *Studies in History and Philosophy of Science*, **16** (1985), pp. 203-246; reprinted in D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity: Einstein Studies* Vol. 1, Boston: Birkhauser, 1989, pp. 5-47.

Norton, John D. (1992) "Einstein, Nordström and the early Demise of Lorentz-covariant, Scalar Theories of Gravitation," *Archive for History of Exact Sciences*, **45**, pp. 17-94.

Norton, John D. (1993) "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute," *Reports on Progress in Physics*, **56**, pp. 791-858.

Norton, John D. (1995) "Mach's Principle before Einstein." pp. 9-57 in J. Barbour and H. Pfister, eds., *Mach's Principle: From Newton's Bucket to Quantum Gravity: Einstein Studies, Vol. 6.* Boston: Birkhäuser.

Norton, John D. (2004) "Einstein's Investigations of Galilean Covariant Electrodynamics prior to 1905," *Archive for History of Exact Sciences*, **59**, pp. 45-105.

Norton, John D. (2006) "Atoms Entropy Quanta: Einstein's Miraculous Argument of 1905," *Studies in History and Philosophy of Modern Physics*, **37**, pp. 71-100.

Renn, Jürgen (2007) *The Genesis of General Relativity*. *Einstein's Zurich Notebook: Introduction and Source.* Volume 1. *Einstein's Zurich Notebook: Commentary and Essays.* Volume 2. *Boston Studies in the Philosophy of Science.* Dordrecht: Springer.

Schulmann, Robert; Kox, A. J.; Janssen, Michel and Illy, József, eds., (1998) *The Collected Papers of Albert Einstein. Volume 8, Part A. The Berlin Years: Correspondence 1914-1917.* Princeton: Princeton University Press.

Stachel, John (1989) "Einstein's Search for General Covariance, 1912-1915," pp. 63-100 in Don Howard and John Stachel, eds., *Einstein and the History of General Relativity*. Boston: Birkhäuser.

Historical and Philosophical Aspects of the Einstein World

Cormac O'Raifeartaigh

This article presents a brief review of some historical and philosophical aspects of Einstein's 1917 paper 'Cosmological Considerations in the General Theory of Relativity', a landmark work that denotes the starting point of modern theoretical cosmology. Our presentation includes a discussion of Einstein's early views of issues such as the relativity of inertia, the curvature of space and the cosmological constant. Particular attention is paid to lesser-known aspects of Einstein's paper such as his failure to test his model against observation, his failure to consider the stability of the model and a slight mathematical confusion concerning the introduction of the cosmological constant term. Taken in conjunction with his later cosmological works, we find that Einstein's approach to cosmology was characterized by a pragmatic search for the simplest model of the universe that was consistent with the principles of relativity and with contemporaneous astronomical observation.

1 Introduction

There is little doubt that Einstein's 1917 paper 'Cosmological Considerations in the General Theory of Relativity' (Einstein 1917a) constituted a key milestone in twentieth century physics. The paper introduced the first relativistic model of the universe, sometimes known as 'Einstein's Static Universe' or the 'Einstein World', marking the starting point of modern theoretical cosmology.

To be sure, a description of the basic physics of the Einstein World can be found in any standard textbook on modern cosmology (Harrison 2000, pp. 355–357; Coles and Lucchin 2002, pp. 26–28). However, while many accounts have been written of

C. O'Raifeartaigh (✉)
School of Science and Computing, Waterford Institute of Technology, Waterford, Ireland
e-mail: coraifeartaigh@wit.ie

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_3

the development of theoretical cosmology from this point onwards, there have been surprisingly few detailed historical analyses of Einstein's 1917 paper itself.[1] The present article presents a discussion of some historical and philosophical aspects of the paper, with an emphasis on lesser-known aspects of the work such as Einstein's failure to test his model against observation, his failure to consider the stability of the model and a slight mathematical confusion concerning the introduction of the cosmological constant term.[2] We also consider Einstein's underlying approach to cosmology in the light of his later cosmological works.

2 Historical Context of the Einstein World

2.1 Biographical Context

Einstein's manuscript 'Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie' or 'Cosmological Considerations in the General Theory of Relativity' (Einstein 1917a) was read to the Prussian Academy of Sciences on February 8th 1917 and published by the Academy on February 15th of that year. Thus the paper, a sizeable ten-page memoir that was to play a seminal role in twentieth century cosmology, appeared only 11 months after the completion of Einstein's greatest and most substantial work, 'Die Grundlage der allgemeinen Relativitätstheorie' or 'The Foundations of the General Theory of Relativity' (Einstein 1916a).[3] The short interval between these two monumental papers is astonishing given that Einstein completed many other works during this period and that he suffered a breakdown in health in early 1917.[4]

On the other hand, it is no surprise from a scientific point of view that Einstein's first foray into cosmology should occur so soon after the completion of the general theory of relativity. After all, it was a fundamental tenet of the general theory that the geometric structure of a region of space-time is not an independent, self-determined entity, but determined by mass-energy (Einstein 1916a). Thus, considerations of the universe at large posed an important test for the new theory. As Einstein later remarked to the Dutch astronomer Willem de Sitter: "For me, though, it was a burning question whether the relativity concept can be followed through to the finish, or whether it leads to contradictions. I am satisfied now that I was able to think the idea through to completion without encountering contradictions" (Einstein

[1]Some exceptions are Kerzberg (1989), Realdi and Peruzzi (2009), and Smeenk (2014).

[2]The present article draws on our comprehensive centenary review of Einstein's paper (O'Raifeartaigh et al. 2017).

[3]The 'Grundlage' paper was submitted to the *Annalen der Physik* on March 20th 1916 and appeared in print on May 11th of that year.

[4]These works included technical papers on quantum theory, gravitational waves, general relativity and a popular book on relativity (O'Raifeartaigh et al. 2017).

1917b). Indeed, it is clear from Einstein's correspondence of 1916 and early 1917 that cosmic considerations—in the sense of the problem of boundary conditions at infinity—were a major preoccupation in the immediate aftermath of the discovery of the covariant field equations (Schulmann et al. 1998, pp. 352–355).

2.2 Cosmology Before 1917

Few quantitative models of the universe were proposed before 1917. One reason was the existence of several puzzles associated with the application of Newton's universal law of gravity to the universe as a whole. For example, it was not clear how a finite Newtonian universe would escape gravitational collapse, as first pointed out by the theologian Richard Bentley, a contemporary of Isaac Newton. Newton's response was to postulate a universe infinite in spatial extent in which the gravitational pull of the stars was cancelled by opposite attractions. However, he was unable to provide a satisfactory answer to Bentley's observation that such an equilibrium would be unstable.[5]

Pioneering work on non-Euclidean geometries in the late nineteenth century led some theoreticians to consider the possibility of a universe of non-Euclidean geometry. For example, Nikolai Lobachevsky considered the case of a universe of hyperbolic (negative) spatial curvature and noted that the lack of astronomical observations of stellar parallax set a minimum value of 4.5 light-years for the radius of curvature of such a universe (Lobachevsky 2010). On the other hand, Carl Friedrich Zöllner noted that a cosmos of spherical curvature might offer a solution to Olbers' paradox[6] and even suggested that the laws of nature might be derived from the dynamical properties of curved space (Zöllner 1872). In the USA, astronomers such as Simon Newcomb and Charles Sanders Peirce took an interest in the concept of a universe of non-Euclidean geometry (Newcomb 1906; Peirce 1891, pp. 174–175), while in Ireland, the astronomer Robert Stawall Ball initiated a program of observations of stellar parallax with the aim of determining the curvature of space (Ball 1881, pp. 92–93; Kragh 2012a). An intriguing theoretical study of universes of non-Euclidean geometry was provided in this period by the German astronomer and theoretician Karl Schwarzschild, who calculated that astronomical observations set a lower bound of 60 and 1500 light-years for the radius of a cosmos of spherical and elliptical geometry, respectively (Schwarzschild 1900). This model was developed further by the German astronomer Paul Harzer, who considered the distribution of stars and the absorption of starlight in a universe of closed geometry (Harzer 1908,

[5]See Norton (1999) and Kragh (2007, pp. 72–74) for a discussion of the Newton–Bentley debate.

[6]This well-known problem concerned the difficulty of reconciling the darkness of the night sky with a universe infinite in space and time (Kragh 2007, pp. 83–86).

pp. 266–267). However, these cosmological considerations had little impact on the physics community and there is no evidence Einstein was aware of them.[7]

The end of the nineteenth century also saw a reconsideration of puzzles associated with Newtonian cosmology in the context of the new concepts of gravitational field and potential.

Defining the gravitational potential Φ as

$$\Phi = G \int \frac{\rho\,(r)}{r} dV \tag{1}$$

where G is Newton's gravitational constant and ρ is the density of matter in a volume V, Newton's law of gravitation could be rewritten in terms of Poisson's equation

$$\nabla^2 \Phi = 4\pi \mathrm{G}\rho \tag{2}$$

where ∇^2 is the Laplacian operator. Distinguished physicists such as Carl Neumann, Hugo von Seeliger and William Thomson noted that the gravitational potential would not be uniquely defined at all distances from a distribution of matter (Neumann 1896, pp. 373–379; Seeliger 1895, 1896; Thomson 1901). Neumann and Seeliger suggested independently that the problem could be solved by replacing Poisson's Eq. (2) with the relation

$$\nabla^2 \Phi - \lambda \Phi = 4\pi G\rho \tag{3}$$

where λ was a decay constant sufficiently small to make the modification significant only at extremely large distances.[8] A different solution to the problem was proposed in 1908 by the Swedish astronomer Carl Charlier, who considered a hierarchical or fractal structure for the universe; in this model the mean density of matter would tend to zero while the density would remain finite in every local location (Charlier 1908). This proposal was later taken up by Franz Selety, who argued that the hierarchic universe could provide a static, Newtonian cosmology alternate to Einstein's relativistic universe (Norton 1999).

[7] See Kragh (2012a, b) for a review of pre-1917 models of the universe of non-Euclidean geometry and their impact.

[8] See North (1965, pp. 17–18) or Norton (1999) for a review of the Neumann–Seeliger proposal.

2.3 Relativistic Cosmology and the Problem of Boundary Conditions at Infinity

In 1915, Einstein published a set of covariant field equations that specified the relation between the geometry of a region of space-time and the distribution of matter/energy within it according to

$$G_{\mu\nu} = -\kappa \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right) \tag{4}$$

where $G_{\mu\nu}$ is a four-dimensional tensor representing the curvature of space-time (known as the Ricci curvature tensor), $T_{\mu\nu}$ is the energy-momentum tensor, T is a scalar and κ is the Einstein constant $8\pi G/c^2$ (Einstein 1915). A description of Einstein's long path to his covariant field equations can be found in reviews such as (Norton 1984; Hoefer 1994; Janssen 2005; Janssen and Renn 2007). As noted in those references, Einstein's thoughts on Mach's principle and the relativity of inertia played a key role in the development of the theory. Indeed, in his well-known 'Prinzipielles' paper of 1918, Einstein explicitly cited the principle as one of three principles[9] fundamental to the development of the theory: "The G-field is completely determined by the masses of the bodies. Since mass and energy are—according to the results of the special theory of relativity—the same, and since energy is formally described by the symmetric energy tensor, it follows that the G-field is caused and determined by the energy tensor of matter" (Einstein 1918a). Further insight into Einstein's understanding of Mach's principle and its relevance to cosmology is offered in the same article: "Mach's Principle (c) is a different story. The necessity to uphold it is by no means shared by all colleagues: but I myself feel it is absolutely necessary to satisfy it. With (c), according to the field equations of gravitation, there can be no G-field without matter. Obviously postulate (c), is closely connected to the space-time structure of the world as a whole, because all masses in the universe will partake in the generation of the G-field" (Einstein 1918a).

Even before the field equations had been published in their final, covariant form, Einstein had obtained an approximate solution for the case of the motion of the planets about the sun (Einstein 1915). In this calculation, the planetary orbits were modelled as motion around a point mass of central symmetry and it was assumed that at an infinite distance from that point, the metric tensor $g_{\mu\nu}$ would revert to flat 'Minkowski' space-time. Indeed, the orbits of the planets were calculated by means of a series of simple deviations from the Minkowski metric. The results corresponded almost exactly with the predictions of Newtonian mechanics with one exception; general relativity predicted an advance of 43 per century in the perihelion

[9]The other principles cited were the principle of relativity and the principle of equivalence (Einstein 1918a).

of the planet Mercury (Einstein 1915). This prediction marked the first success of the general theory, as the anomalous behaviour of Mercury had been well-known to astronomers for some years but had remained unexplained in Newton's theory. The result was a source of great satisfaction to Einstein and a strong indicator that his new theory of gravity was on the right track (Earman and Janssen 1993).

In early 1916, Karl Schwarzschild obtained the first exact solution to the general field equations, again pertaining to the case of a mass point of central symmetry (Schwarzschild 1916). Einstein was surprised and delighted by the solution, declaring in a letter to Schwarzschild in January 1916 that "I would not have expected that the exact solution to the problem could be formulated so simply" (Einstein 1916b). In the Schwarzschild solution, it was once again assumed that sufficiently far from a material body, the space-time metric would revert to flat space-time. The imposition of such 'boundary conditions' was not unusual in field theory; however, such an approach could hardly be applied to the universe as a whole, as it raised the question of the existence a privileged frame of reference at infinity. Moreover, the assumption of a Minkowski metric an infinite distance away from matter was in obvious conflict with Einstein's understanding of Mach's principle.

Einstein's correspondence suggests that he continued to muse on the problem of boundary conditions at infinity throughout the year 1916. For example, a letter written to his old friend Michele Besso in May 1916 contains a reference to the problem, as well as an intriguing portend of Einstein's eventual solution: "In gravitation, I am now looking for the boundary conditions at infinity; it certainly is interesting to consider to what extent a finite world exists, that is, a world of naturally measured finite extension in which all inertia is truly relative" (Einstein 1916c). In the autumn of 1916, Einstein visited Leiden in Holland for a period of 3 weeks. There he spent many happy hours discussing his new theory of gravitation with his great friends Henrik Lorentz and Paul Ehrenfest. Also present at these meetings was the Dutch astronomer and theorist Willem de Sitter. A number of letters and papers written shortly afterwards by de Sitter (1916a, b) suggest that many of these discussions concerned the problem of boundary conditions, i.e., the difficulty of finding boundary conditions at infinity that were consistent with the principle of relativity and with Mach's principle. In one such article, de Sitter gives evidence that, at this stage, Einstein's solution was to suggest that, at an infinite distance from gravitational sources, the components of the metric tensor [$g_{\mu\nu}$] would reduce to degenerate values: "Einstein has, however, pointed out a set of degenerated g_{ij} which are actually invariant for all transformations in which, at infinity x_4 is a pure function of x'_4. They are:

$$\begin{pmatrix} 0 & 0 & 0 & \infty \\ 0 & 0 & 0 & \infty \\ 0 & 0 & 0 & \infty \\ \infty & \infty & \infty & \infty^2 \end{pmatrix}$$

These are then the "natural values, and any deviation from them must be due to material sources....At very large distances from all matter the g_{ij} would gradually converge towards the degenerated values" (de Sitter 1916a).

However, de Sitter highlights a potential flaw in Einstein's proposal. Since observation of the most distant stars showed no evidence of spatial curvature, it was puzzling how the 'local' Minkowskian values of the gravitational potentials $g_{\mu\nu}$ arose from the postulated degenerate values at infinity. According to de Sitter, Einstein proposed that this effect was due to the influence of distant masses: "Now it is certain that, in many systems of reference (i.e., in all Galilean systems) the g_{ij} at large distances from all material bodies known to us actually have the [Minkowski] values. On Einstein's hypothesis, these are special values which, since they differ from [degenerate] values, must be produced by some material bodies. Consequently there must exist, at still larger distances, certain unknown masses which are the source of the [Minkowski] values, i.e., of all inertia" (de Sitter 1916a). Yet no trace of such masses were observable by astronomy: "We must insist on the impossibility that any of the known fixed stars or nebulae can form part of these hypothetical masses. The light even from the farthest stars and nebulae has approximately the same wavelength as light produced by terrestrial sources. ...the deviation of the g_{ij} from the Galilean values ... is of the same order as here, and they must therefore be still inside the limiting envelope which separates our universe from the outer parts of space, where the g_{ij} have the [degenerate] values." Indeed, de Sitter concludes that the hypothetical distant masses essentially play the role of absolute space in classical theory. "If we believe in the existence of these supernatural masses, which control the whole physical universe without having ever being observed then the temptation must be very great indeed to give preference to a system of co-ordinates relatively to which they are at rest, and to distinguish it by a special name, such as "inertial system" or "ether." Formally the principle of relativity would remain true, but as a matter of fact we would have returned to the absolute space under another name" (de Sitter 1916a).

Einstein and de Sitter debated the issue of boundary conditions at infinity in correspondence for some months. A review of their fascinating debate can be found in references such as (Kerzberg 1989; Hoefer 1994; Schulmann et al. 1998, pp. 353–354; Realdi and Peruzzi 2009). We note here that Einstein conceded defeat on the issue in a letter written to de Sitter on November 4th 1916: "I am sorry for having placed too much emphasis on the boundary conditions in our discussions. This is purely a matter of taste which will never gain scientific significance. . . . Now that the covariant field equations have been found, no motive remains to place such great weight on the total relativity of inertia" (Einstein 1916d). However, the closing paragraph of the same letter indicates that Einstein had not completely given up on the notion of the relativity of inertia: "On the other hand, you must not scold me for being curious enough still to ask: Can I imagine a universe or the universe in such a way that inertia stems entirely from the masses and not at all from the boundary conditions? As long as I am aware that this whim does not touch the core of the theory, it is innocent; by no means do I expect you to share this curiosity" (Einstein 1916d). Notice of a successful conclusion to Einstein's quest appears in another letter to de Sitter, written on 2nd February 1917: "Presently I am writing a paper on the boundary conditions in gravitation theory. I have completely abandoned my idea on the degeneration of the $g_{\mu\nu}$, which you rightly disputed. I am curious to see what you will say about the rather outlandish conception I have now set my sights on" (Einstein 1917c). The 'outlandish conception' was the postulate of a universe of closed spatial geometry, as described below.

3 Einstein's 1917 Paper

A surprising feature of Einstein's 1917 cosmological memoir is the sizeable portion of the paper concerned with Newtonian cosmology. This analysis had two important aims. In the first instance, Einstein was no doubt pleased to show that his new theory of gravitation could overcome a well-known puzzle associated with Newtonian cosmology. Second, a suggested ad hoc modification of Newtonian gravity provided a useful analogy for a necessary modification of the field equations of relativity.

Einstein's assault on Newtonian cosmology is two-pronged. First he establishes from symmetry principles that Newtonian gravity only allows for a finite island of stars in infinite space. Then he suggests from a consideration of statistical mechanics that such an island would evaporate, in contradiction with the presumed static nature of the universe. His solution to the paradox is the introduction of a new term to Poisson's equation. This solution is very similar to that of Seeliger and Neumann, although Einstein was not aware of this work at the time (O'Raifeartaigh et al. 2017). A year later, Einstein presented a simpler argument against the Newtonian universe in terms of lines of force; this argument was published in the third edition of his popular book on relativity (Einstein 1918b, p. 123) and retained in all later editions of the book.

We note that a few years after the publication of the 1917 memoir, the Austrian physicist Franz Selety noted that the hierarchic cosmology proposed by Carl Charlier (above) avoided the paradox identified by Einstein (Selety 1922). Einstein conceded the point, but objected to the Charlier's model on the grounds that it was anti-Machian (Einstein 1922b).[10]

3.1 On the Basic Assumptions of Einstein's Model

It is clear from Einstein's 1917 memoir that the starting point of his cosmic model was the assumption of a universe with a static distribution of matter, uniformly distributed over the largest scales and of non-zero average density. Considering the issue of stasis first, Einstein argued for a quasi-static distribution of matter based

[10] See Norton (1999) for a discussion of the Einstein–Selety debate.

on the small velocities of the stars: “The most important fact that we draw from experience as to the distribution of matter is that the relative velocities of the stars are very small as compared with the velocity of light. So ....there is a system of reference relatively to which matter may be looked upon as being permanently at rest” (Einstein 1917a). It is generally agreed amongst historians and physicists that this assumption was reasonable at the time (Hoefer 1994; Kragh 2007, pp. 131–132; Nussbaumer and Bieri 2009, pp. 72–76). There is no evidence that Einstein was aware at this time of Slipher’s observations of the redshift of light from the spiral nebulae, while the extra-galactic nature of the spirals had yet to be established. Indeed, many years were to elapse before the demonstration of a linear relation between the redshifts of the distant galaxies and their distance (Hubble 1929), the first evidence for a non-static universe. However, it is worth noting that Einstein’s stellar argument was questioned by de Sitter: “We only have a snapshot of the world, and we cannot and must not conclude from the fact that we do not see any large changes on this photograph that everything will always remain as at that instant when the picture was taken.” (de Sitter 1917b). It could also be argued that Einstein erred philosophically in inferring *global* stasis from astronomical observations of the *local* environment (Kerzberg 1989; Smeenk 2014, p. 241); however, we find his assumption reasonable in the context of the widespread contemporaneous belief that the universe was not much larger than the Milky Way.

It is sometimes stated that Einstein’s assumption of stasis prevented him from predicting the expansion of the universe many years before the phenomenon was discovered by astronomers. This statement may be true in a literal sense, but we find it somewhat anachronistic. It is clear throughout his cosmological memoir that Einstein’s interest lay in establishing whether he could achieve a description of the universe consistent with the foundational principles of general relativity (including, in particular, Mach’s principle) and with astronomical observation. Thus, the exploration of solutions to the field equations for the case of a non-static cosmos would have been of little interest to him in 1917. Many years later, Einstein stated that the assumption of a static universe “appeared unavoidable to me at the time, since I thought that one would get into bottomless speculations if one departed from it” (Einstein 1945, p. 137). Indeed, it could be argued that the common moniker ‘Einstein’s static model of the universe’ is a little misleading, as it implies a choice from a smorgasbord of possible models of the known universe. Historically speaking, a more accurate title would be ‘Einstein’s model of the Static Universe’.

In some ways, Einstein’s assumption of matter ‘as being uniformly distributed over enormous spaces’ was more radical than his assumption of stasis. Technically speaking, this assumption implied a universe that was both isotropic and homogeneous, at least on the largest scales, an assumption that was at odds with astronomical observations. Thus, the assumption was more of an assumed principle and indeed it was later named the ‘Cosmological Principle’ (Milne 1935, p. 24). One reason for the principle was its undoubted simplicity, as the assumption of homogeneity and isotropy greatly simplified the business of solving the field equations. A deeper reason may have been that the Cosmological Principle chimed with a Copernican approach to cosmology and with the spirit of relativity (Bondi 1952, pp. 11–13). After all, to assume a universe with a non-uniform distribution of matter on the largest scales was to assume a universe in which all viewpoints were not equivalent, in contradiction with basic tenets of relativity (Milne 1933).

3.2 On Spatial Curvature

In his 1917 memoir, Einstein's solution to the problem of boundary conditions at infinity was to banish the boundaries by postulating a world of closed, spherical spatial curvature. In this manner, the Einstein World explicitly incorporated his view of the relativity of inertia.[11] It was later shown that closed geometry was the *only* possibility for a universe with a static, homogeneous distribution of matter of non-zero average density. Thus, Einstein's view of Mach's principle was a useful, but not strictly necessary, guide to his first model of the universe, just as it was a guide on his path to the field equations.

Following the publication of the 1917 paper, colleagues such as Erwin Freundlich, Felix Klein and Willem de Sitter suggested in correspondence to Einstein that elliptical geometry would also satisfy the requirements of his cosmology (O'Raifeartaigh et al. 2017). Einstein quickly conceded the point, noting that the relation between the radius of curvature and the mean density of matter remained unchanged. For example, he remarked to Klein: "As I have never done non-Euclidean geometry, the more obvious elliptical geometry had escaped me . . . .my observations are just altered thus, that the space is half as large; the relation between R (the radius of curvature) and ρ (mean density of matter) is retained" (Einstein 1917d). A few months later, he commented to de Sitter: "When I was writing the paper, I did not yet know about the elliptical possibility . . . ..this possibility seems more likely to me as well" (Einstein 1917e). This preference was cited by de Sitter in his classic paper of 1917: "The elliptical space is, however, really the simpler case, and it is preferable to adopt this for the physical world" . . . this is also the opinion of Einstein" (de Sitter 1917a). Neither Einstein nor de Sitter made clear in their correspondence why they prefer elliptical geometry; one explanation may be that they viewed this geometry as more general than spherical.

[11] Unfortunately, Einstein's expression "*räumlich geschlossen*" or "spatially closed" is mistranslated throughout the official English translation of the paper as "spatially finite" (Einstein 1917a; O'Raifeartaigh et al. 2017).

3.3 On the Cosmological Constant

In his cosmological memoir, Einstein soon found that the hypothesis of closed spatial geometry was not sufficient to achieve a successful relativistic model of the universe. Instead, a consistent solution could only be achieved with the introduction of an additional term $\lambda g_{\mu\nu}$ to the field equations, according to

$$G_{\mu\nu} - \lambda g_{\mu\nu} = -\kappa \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right) \tag{5}$$

where λ was a universal constant that became known as the *cosmological constant*. Einstein then showed that the modified field Eq. (5) has the solution

$$\lambda = \frac{\kappa\rho}{2} = \frac{1}{R^{2\cdot}} \tag{6}$$

where ρ and R represent the mean density of matter and the radius of the cosmos, respectively (Einstein 1917a). In this manner, Einstein's 1917 model of the cosmos gave an apparently satisfactory relation between the size of the universe and the amount of matter it contained.

Thus, Einstein's model appears to have evolved according to the following sequence of assumptions: uniform, static distribution of matter → closed spatial geometry → introduction of additional term to the field equations. While the general theory allowed such a modification of the field equations, Einstein seems to have anticipated some resistance to the term; it is interesting that he forewarns the reader of what is to come on three separate occasions in the paper. Indeed, it could be argued that much of Einstein's 1917 memoir can be read as a lengthy justification for the introduction of the cosmological constant term to relativity!

Some historians have found Einstein's use of the cosmological constant term in his 1917 memoir somewhat ambiguous and argue that his view of the term wavers throughout the paper (Kerzberg 1989). In our view, the *purpose* of the term is clear throughout the paper, both in the stated text and in the underlying physics of the model, and is summarized quite precisely in the final sentence: "That term is necessary only for the purpose of making possible a quasi-static distribution of matter, as required by the fact of the small velocities of the stars." That said, there is little doubt that the cosmological constant term posed a significant challenge to Einstein in terms of interpretation. Indeed, it is striking that no interpretation of the physics underlying the term is presented anywhere in the 1917 paper and there is ample evidence in Einstein's later writings that he viewed his modification of the field equations as an uncomfortable mathematical necessity. For example, in March 1917, Einstein remarked to Felix Klein: "The new version of the theory means, formally, a complication of the foundations and will probably be looked upon by almost all our colleagues as an interesting, though mischievous and superfluous stunt, particularly since it is unlikely that empirical support will be obtainable in the foreseeable future. But I see the matter as a necessary addition, without which neither inertia nor geometry are truly relative" (Einstein 1917d). Similarly, when de Sitter commented in a letter of March 20th: "I personally much prefer the four-dimensional system, but even more so the original theory, without the undeterminable λ, which is just philosophically and physically desirable (de Sitter 1917c), Einstein responded: "In any case, one thing stands. The general theory of relativity allows the addition of the term $\lambda g_{\mu\nu}$ in the field equations. One day, our actual knowledge of the composition of the fixed-star sky, the apparent motions of fixed stars, and the position of spectral lines as a function of distance, will probably have come far enough for us to be able to decide empirically the question of whether or not λ vanishes. Conviction is a good mainspring, but a bad judge!" (Einstein 1917f).

In March 1918, the Austrian physicist Erwin Schrödinger suggested that a consistent model of a static, matter-filled cosmos could be obtained from Einstein's field equations without the introduction of the cosmological constant term (Schrödinger 1918). Essentially, Schrödinger's proposal was that Einstein's solution could be obtained from the unmodified field Eq. (4) if a negative-pressure term was added to the 'source' tensor on the right-hand side of the equations, i.e., by replacing Einstein's energy-momentum tensor by the tensor

$$T^{\mu\nu} = \begin{pmatrix} -p & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & \rho - p \end{pmatrix} \tag{7}$$

where ρ is the mean density of matter and p is the pressure (defined as $p = \lambda/\kappa$).

Einstein's response was that Schrödinger's formulation was entirely equivalent to that of his 1917 memoir, provided the negative-pressure term was constant (Einstein 1918c).[12] This response seems at first surprising; Schrödinger's new term may have been *mathematically* equivalent to that of Einstein's but the underlying physics was surely different. However, in the same paper, Einstein gave his first physical interpretation of the cosmological term, namely that of a negative mass density: "In terms of the Newtonian theory . . . a modification of the theory is required such that "empty space" takes the role of gravitating negative masses which are distributed all over the interstellar space" (Einstein 1918c).

Within a year, Einstein proposed a slightly different interpretation of the cosmological constant term. Rewriting the field equations in a slightly different format, he opined that the cosmological constant now took the form of a constant of integration, rather than a universal constant associated with cosmology: "But the new formulation has this great advantage, that the quantity appears in the fundamental equations as a constant of integration, and no longer as a universal

[12]Schrödinger also suggested that the pressure term might be time variant, anticipating the modern concept of quintessence, but this suggestion was too speculative for Einstein (1918c).

constant peculiar to the fundamental law" (Einstein 1919). Indeed, a letter to Michele Besso suggests that Einstein had arrived at a similar interpretation a year earlier using a variational principle (Einstein 1918d). A follow-up letter to Besso suggests that at one point, Einstein considered the two views to be equivalent: "Since the world exists as a single specimen, it is essentially the same whether a constant is given the form of one belonging to the natural laws or the form of an 'integration constant'" (Einstein 1918e).

Thus, there is little doubt that a satisfactory interpretation of the physics of the cosmological constant term posed a challenge for Einstein in these years. One contributing factor to this ambiguity may be a slight mathematical confusion concerning manner in which the term was introduced. As several scholars have noted (Norton 1999; Harvey and Schucking 2000), Einstein's modification of the field equations in his memoir was not in fact exactly analogous to his modification of Newtonian gravity, as he claimed, i.e., the modified field Eq. (5) do not reduce in the Newtonian limit to the modified Poisson Eq. (3), but to the slightly different relation

$$\nabla^2\phi + c^2\,\lambda = 4\pi \mathrm{G}\rho \tag{8}$$

This might seem a rather pedantic point, but the error may have been significant with regard to Einstein's interpretation of the term. Where he intended to introduce a term to the field equations representing an attenuation of the gravitational interaction at large distances, he in fact introduced a term representing a tendency for empty space to expand, a concept that would have been in conflict with his view of Mach's principle at the time.

3.4 *On Testing the Model Against Observation*

A curious aspect of Einstein's 1917 memoir is that, having established a pleasing relation between the geometry of the universe and the matter it contained, he made no attempt to test the model against empirical observation. After all, even a rough estimate of the mean density of matter ρ in Eq. (6) would give a value for the cosmic radius R and the cosmological constant λ. These values could then have been checked against observation; one could expect an estimate for R that was not smaller than astronomical estimates of the size of the distance to the furthest stars, and an estimate for λ that was not too large to be compatible with observations of the orbits of the planets. No such calculation is to be found in the 1917 memoir. Instead, Einstein merely declares at the end of the paper that the model is logically consistent: "At any rate, this view is logically consistent, and from the standpoint of the general theory of relativity lies nearest at hand; whether, from the standpoint of present astronomical knowledge, it is tenable, will not here be discussed."

We have previously noted that Einstein did attempt such a calculation in his correspondence around this time (O'Raifeartaigh et al. 2017). Taking a value of $\rho = 10^{-22}$ g/cm^3 for the mean density of matter,[13] he obtained from Eq. (6) an estimate of 10^7 light-years for the radius of his universe, a result he found unrealistic. As he stated in a letter to Paul Ehrenfest: "From the measured stellar densities, a universe radius of the order of magnitude of 10^7 light-years results, thus unfortunately being very large against the distances of observable stars" (Einstein 1917g). This comment implies that, like many of his contemporaries at the time, Einstein did not believe that the universe was significantly larger than the Milky Way. However, Einstein does not appear to have taken such calculations too seriously, presumably because he lacked confidence in astronomical estimates of the mean density of matter. As he remarked in a letter to Erwin Freundlich: "..The matter of great interest here is that not only R but also ρ must be individually determinable astronomically, the latter quantity at least to a very rough approximation, and then my relation between them ought to hold. Maybe the chasm between the 10^4 and 10^7 light years can be bridged after all. That would mean the beginning of an epoch in astronomy" (Einstein 1917h). Later writings also suggest that Einstein viewed the average density of matter in the universe as an unknown quantity (Einstein 1921; O'Raifeartaigh et al. 2017).

3.5 On the Stability of the Einstein World

Perhaps the strangest aspect of Einstein's 1917 memoir is his failure to consider the stability of his cosmic model. After all, Eq. (6) drew a direct equation between a universal constant λ, the radius of the universe R, and the density of matter ρ. But the quantity ρ represented a *mean* value for the density of matter, arising from the theoretical assumption of a uniform distribution of matter on the largest scales. In the real universe, one would expect a natural variation in this parameter in time and space, raising the question of the stability of the model against such perturbations. It was later shown that the Einstein World is generally unstable against such density perturbations: instead of oscillating around a stable solution, a slight increase in the density of matter (without a corresponding change in λ) would cause the universe to contract, become more dense and contract further, while a slight decrease in density would result in a runaway expansion (Eddington 1930).[14]

It is curious that Einstein did not consider this aspect of his model in 1917; some years later, it was a major reason for his rejection of the model, as described in the next section.

[13]Einstein does not give a reference for his estimate of the mean density of matter in his correspondence but it is in reasonable agreement with that given by de Sitter (1917a).

[14]See Gibbons (1987) for further discussion of the stability of the Einstein World.

4 The Einstein–de Sitter Debate

In July 1917, Willem de Sitter published a paper in which he noted that the modified field equations allowed a cosmological solution for the case of a universe with no matter content (de Sitter 1917a). In this cosmology, Einstein's matter-filled three-dimensional universe of spherical spatial geometry was replaced by an empty four-dimensional universe of closed *space-time* geometry. It should come as no surprise that Einstein was greatly perturbed by de Sitter's solution, as the model was in direct conflict with his understanding of Mach's principle in these years. A long debate ensued between the two physicists concerning the relative merits of the two models that has been extensively described in the literature.[15] Eventually, Einstein made his criticisms public in a paper of 1918: "It appears to me that one can raise a grave argument against the admissibility of this solution.....In my opinion, the general theory of relativity is a satisfying system only if it shows that the physical qualities of space are completely determined by matter alone. Therefore no $g_{\mu\nu}$- field must exist (that is no space-time continuum is possible) without matter that generates it" (Einstein 1918f). Einstein also raised a technical objection to de Sitter's model, namely that it appeared to contain a space-time singularity. In the years that followed, Einstein continued to debate the de Sitter model with physicists such as Kornel Lanczos, Hermann Weyl, Felix Klein and Gustav Mie. Throughout this debate, Einstein did not waver from his core belief that a satisfactory cosmology should describe a universe that was globally static with a metric structure that was fully determined by matter.[16] Einstein eventually conceded that the apparent singularity in the de Sitter universe was an artefact of co-ordinate representation (Einstein 1918g), but he never formally retracted his criticism of the de Sitter universe in the literature, nor did he refer to the de Sitter model in his formal writings on cosmology in these years (O'Raifeartaigh et al. 2017).

5 Einstein and the Expanding Universe

In 1922, the Russian physicist Alexander Friedman suggested that non-static solutions of the Einstein field equations should be considered in relativistic models of the cosmos (Friedman 1922). Einstein publicly faulted Friedman's analysis on the basis that it contained a mathematical error (Einstein 1922a). When it transpired that the error lay in Einstein's criticism, it was duly retracted (Einstein 1923a). However, an unpublished draft of Einstein's retraction demonstrates that he did not consider

[15] See for example Kerzberg (1989), Schulmann et al. (1998, pp. 352–354), Realdi and Peruzzi (2009).

[16] See Schulmann et al. (1998, pp. 355–357) for a discussion of the Einstein–deSitter–Weyl–Klein debate.

Friedman's cosmology to be realistic: "to this a physical significance can hardly be ascribed" (Einstein 1923b).[17]

A few years later, the Belgian physicist Georges Lemaître independently derived time-varying equations for the radius of the cosmos from Einstein's modified field equations. Aware of Slipher's observations of the redshifts of the spiral nebulae, and of emerging measurements of the distance of the spirals by Edwin Hubble, Lemaître suggested that the recession of the nebulae was a manifestation of the expansion of space from a pre-existing Einstein World of cosmic radius $R_0 = 1/\sqrt{\lambda}$ (Lemaître 1927). This work was brought to Einstein's attention by Lemaître himself, only to have expanding cosmologies dismissed as *'abominable'*. According to Lemaître, Einstein's rejection probably stemmed from a lack of knowledge of developments in astronomy: "Je parlais de vitesses des nébeleuses et j'eus l'impression que Einstein n'était guère au courant des faits astronomiques" (Lemaître 1958).

In 1929, Edwin Hubble published empirical evidence of a linear relation between the redshifts of the spiral nebulae and their radial distance (Hubble 1929).[18] Many theorists interpreted the observations in terms of a relativistic expansion of space, and a number of cosmic models of the Friedman–Lemaître type were advanced for diverse values of cosmic parameters. Einstein himself overcame his earlier distrust of expanding models of the cosmos, stating during a sojourn at the California Institute of Technology in 1931:"New observations by Hubble and Humason concerning the redshift of light in distant nebulae make the presumptions near that the general structure of the universe is not static" (AP 1931a) and "The redshift of the distant nebulae have smashed my old construction like a hammer blow" (AP 1931b). A recently discovered manuscript indicates that Einstein first considered a steady-state model of the universe on learning of Hubble's observations; however, the model led to a null solution and he quickly abandoned the attempt (O'Raifeartaigh et al. 2014; Nussbaumer 2014a). In April 1931, Einstein published a model of the expanding cosmos based on Friedman's (1922) analysis, with the cosmological term removed, deriving simple expressions relating the rate of cosmic expansion (an observable that could be measured from the recession of the nebulae) to the radius of the cosmos, the density of matter and the timespan of the expansion.[19] It is interesting to note that Einstein provided a two-fold justification for abandoning the cosmological constant term in this paper. In the first instance, the term was unsatisfactory because it did not provide a stable solution: "It can also be shown ... that this solution is not stable. On these grounds alone, I am no longer inclined to ascribe a physical meaning to my former solution" (Einstein 1931). In the second instance, the term was unnecessary because the assumption of stasis was no longer justified by observation: "Now that it has become clear from Hubbel's [sic] results that the extra-galactic nebulae

[17] A detailed account of this episode can be found in Nussbaumer and Bieri (2009, pp. 91–92).

[18] Although Lemaître had derived such a relation in 1927 from theory, the empirical verification of the relation is attributable to Hubble (O'Raifeartaigh 2013; Kragh 2018).

[19] We have recently provided an analysis and first English translation of this paper (O'Raifeartaigh and McCann 2014) and noted that Einstein's calculations contain a systematic error.

are uniformly distributed throughout space and are in dilatory motion (at least if their systematic redshifts are to be interpreted as Doppler effects), assumption (2) concerning the static nature of space has no longer any justification" (Einstein 1931). A year later, Einstein proposed an even simpler model of the expanding universe in conjunction with de Sitter; in this model, both the cosmological constant and spatial curvature were removed (Einstein and de Sitter 1932).

Thus it is clear that, when presented with empirical evidence for a dynamic universe, Einstein lost little time in abandoning his static cosmology.[20] He also abandoned the cosmological constant term and was never to re-instate it in his cosmological models. Indeed, he is reputed to have described the term in later years as his 'biggest blunder'. Whether Einstein used these exact words has been the subject of some debate,[21] but his considered view of the cosmological constant term was made clear in a 1945 review of relativistic cosmology: "If Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmologic member would never have been introduced. It seems now so much less justified to introduce such a member into the field equations, since its introduction loses its sole original justification—that of leading to a natural solution of the cosmologic problem" (Einstein 1945, p. 130). This passage neatly encapsulates Einstein's matter-of-fact approach to cosmology—if the known universe could be modelled without the cosmological constant term, why include it?

6 Conclusions

In his 1917 cosmological memoir, Einstein demonstrated that his newly minted general theory of relativity could give a model of the universe that was consistent with the founding principles of the theory, including Mach's principle, and with astronomical observation. The price was the hypothesis of closed spatial geometry for the cosmos and a modification of the field equations of general relativity. A slight mathematical inaccuracy associated with Einstein's introduction of the cosmological constant is intriguing; it is possible that this ambiguity may have affected his interpretation of the term. It is also interesting that Einstein made no formal attempt to test his model against empirical observation; later writings suggest that he distrusted astronomical estimates of the mean density of matter in the universe. Perhaps the most curious aspect of Einstein's 1917 memoir is his failure to consider the stability of his cosmic model. When he formally abandoned the Einstein World in 1931, it was on the twin grounds that the model was both theoretically unstable and in conflict with empirical observation.

We note finally that the Einstein World has become a topic of renewed interest in today's cosmology. Some theorists have become interested in the hypothesis of a

[20] See Nussbaumer (2014b) for further details on Einstein's conversion to expanding cosmologies.

[21] We have recently provided an interrogation of this story (O'Raifeartaigh and Mitton 2018).

universe that expands from a static Einstein World after an indefinite period of time, thus reviving Lemaître's 1927 model in the context of the modern theory of cosmic inflation. It is thought that this scenario, known as 'the emergent universe', might be useful in addressing major difficulties in modern cosmology such as the horizon problem, the quantum gravity era and the initial singularity.[22] Whether the emergent universe will offer a plausible, consistent description of the origins and evolution of our universe is not yet known, but we note, as so often, the relevance of past models of the universe in today's research.

Acknowledgements The author acknowledges the use of The Collected Papers of Albert Einstein (CPAE), an invaluable online archive of Einstein's original works complete with annotations and editorial comments, provided by Princeton University Press in conjunction with the California Institute of Technology and the Hebrew University of Jerusalem. Helpful comments from the editors and an anonymous referee are also acknowledged.

References

Associated Press Report. 1931a. Prof. Einstein begins his work at Mt. Wilson. *New York Times*, Jan 3, p1

Associated Press Report. 1931b. Red shift of nebulae a puzzle, says Einstein. *New York Times*, Feb 12, p 2

Ball, R.S. 1881. The distances of stars. *Nature* **24**: 91-92

Barrow, J., Ellis G.F.R, Maartens R. and C.G. Tsagas 2003. On the stability of the Einstein static universe. *Class. Quant. Gravity*. **20** (11): L155-161

Bondi, H. 1952. *Cosmology*. Cambridge University Press, Cambridge

Charlier, C. 1908. Wie eine unendliche Welt aufgebaut sein kann. *Arkiv Mat. Ast. Fysik* **4**: 1-15

Coles, P and F. Lucchin. 2002. *Cosmology: The Origin and Evolution of Cosmic Structure*. Wiley & Sons, Chichester

de Sitter, W. 1916a. On Einstein's theory of gravitation and its astronomical consequences. Second Paper. MNRAS **77:** 155-184

de Sitter, W 1916b. De relativiteit der rotatie in de theorie van Einstein. *Kon. Akad. van Wetensch. Amsterdam* **25**: 499-504. Or 'On the relativity of rotation in Einstein's theory.' *Proc. Roy. Neth. Acad. Art. Sci.* **19** (1): 527-532 (1917)

de Sitter, W. 1917a. On Einstein's theory of gravitation and its astronomical consequences. Third paper. *MNRAS* **78:** 3-28

de Sitter, W. 1917b. Letter to Albert Einstein, April 1st. CPAE 8 (Doc. 321).

de Sitter, W. 1917c. Letter to Albert Einstein, March 20th. CPAE 8 (Doc. 313).

Earman, J. and M. Janssen 1993. Einstein's Explanation of the Motion of Mercury's Perihelion. In *The Attraction of Gravitation: New Studies in the History of General Relativity*. J. Earman et al. (eds.) Birkhaeuser pp 129-159

Eddington, A.S. 1930. On the instability of Einstein's spherical world. *MNRAS* **90**: 668-678

Einstein, A. 1915. Die Feldgleichungen der Gravitation. *Sitz. König. Preuss. Akad*. 844-847. Or 'The field equations of gravitation' CPAE **6** (Doc. 25).

Einstein, A. 1916a. Die Grundlage der allgemeinen Relativitätstheorie. *Ann. Physik*. **49:** 769-822. Or 'The foundation of the general theory of relativity' CPAE **6** (Doc. 30)

[22] See Barrow et al. (2003) for an introduction to emergent cosmology.

Einstein, A. 1916b. Letter to Karl Schwarzschild, January 9th. CPAE **8** (Doc. 181).
Einstein, A. 1916c. Letter to Michele Besso, May 14th. CPAE **8** (Doc. 219).
Einstein, A. 1916d. Letter to Willem de Sitter, November 4th. CPAE **8** (Doc. 273).
Einstein, A. 1917a. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitz. König. Preuss. Akad.* 142-152. Or 'Cosmological considerations in the general theory of relativity' CPAE **6** (Doc. 43).
Einstein, A. 1917b. Letter to Willem de Sitter, March 12th. CPAE **8** (Doc. 311).
Einstein, A. 1917c. Letter to Willem de Sitter, February 2nd. CPAE **8** (Doc. 293).
Einstein, A. 1917d. Letter to Felix Klein, March 26th. CPAE **8** (Doc. 319).
Einstein, A. 1917e. Letter to Willem de Sitter, June 28th. CPAE **8** (Doc. 359).
Einstein, A. 1917f. Letter to Willem de Sitter, April 14th. CPAE **8** (Doc. 325).
Einstein, A. 1917g. Letter to Paul Ehrenfest, February 14th. CPAE **8** (Doc. 298).
Einstein, A. 1917h. Letter to Erwin Freundlich, February 18th. CPAE **8** (Doc. 300).
Einstein, A. 1918a. Prinzipielles zur allgemeinen Relativitätstheorie. *Annal. Physik***55:** 241-244. Or 'On the foundations of the general theory of relativity' CPAE **7** (Doc. 4).
Einstein 1918b. *Über die Spezielle und die Allgemeine Relativitätstheorie*. Vieweg (Braunschweig). 3rd Edition. CPAE **6** (Doc.42).
Einstein, A. 1918c. Bemerkung zu Herrn Schrödingers Notiz "Über ein Lösungssystem
der allgemein kovarianten Gravitationsgleichungen. *Phys. Zeit.***19:** 165-166. Or 'Comment on Schrödinger's Note "On a system of solutions for the generally covariant gravitational field equations"' CPAE **7** (Doc. 3).
Einstein 1918d. Letter to Michele Besso, July 29th. CPAE **8** (Doc. 591).
Einstein 1918e. Letter to Michele Besso, August 20th. CPAE **8** (Doc. 604).
Einstein, A. 1918f. Kritisches zu einer von Hrn. De Sitter gegebenen Lösung der Gravitationsgleichungen. *Sitz. König. Preuss. Akad.* 270-272. Or 'Critical comment on a solution of the gravitational field equations given by Mr. de Sitter' CPAE **7** (Doc. 5).
Einstein 1918g. Letter to Felix Klein, June 20th. CPAE 8 (Doc. 567).
Einstein, A. 1919. Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle? *Sitz. König. Preuss. Akad.* 349-356. Or 'Do gravitation fields play an essential part in the structure of the elementary particles of matter?' CPAE **7** (Doc. 17).
Einstein, A. 1921. *Geometrie und Erfahrung*. Springer, Berlin. Or 'Geometry and Experience'. CPAE **7** (Doc. 52).
Einstein, A. 1922a. Bemerkung zu der Arbeit von A. Friedmann "Über die Krümmung des Raumes" *Zeit. Phys.***11**: 326. Or 'Comment on A. Friedmann's paper "On The Curvature of Space"' CPAE **13** (Doc. 340).
Einstein, A. 1922b. Bemerkung zu der Franz Seletyschen Arbeit "Beiträge zum kosmologischen System". *Ann. Physik***69**: 436–438. Or 'Comment on Franz Selety's paper "Contributions to the Cosmological System"' CPAE **13** (Doc. 370).
Einstein, A. 1923a. Notiz zu der Arbeit von A. Friedmann "Über die Krümmung des Raumes" *Zeit. Phys.***16**: 228. Or 'Note to the paper by A. Friedmann "On the Curvature of Space"' CPAE **14** (Doc. 51).
Einstein, A. 1923b. Notiz zu der Arbeit von A. Friedmann "Über die Krümmung des Raumes" . The Albert Einstein Archives. Doc. **1**-26.
Einstein, A. 1931. Zum kosmologischen Problem der allgemeinen Relativitätstheorie. *Sitz. König. Preuss. Akad.* 235-237
Einstein, A. 1945. On the 'cosmologic problem'. Appendix I to *The Meaning of Relativity*. Princeton University Press, Princeton (3rd Ed.) 112-135
Einstein, A. and W. de Sitter. 1932. On the relation between the expansion and the mean density of the universe. *PNAS***18**(3): 213-214
Friedman, A. 1922. Über die Krümmung des Raumes. *Zeit. Physik.***10**: 377-386. Or 'On the curvature of space' *Gen. Rel. Grav.***31** (12): 1991 (1999)
Gibbons, G. 1987. The entropy and stability of the universe. *Nuc. Phys. B***292**: 784-792
Harrison, E. 2000. *Cosmology*. 2nd Edition. Cambridge University Press, Cambridge

Harvey, A. and E. Schucking 2000. Einstein's mistake and the cosmological constant. *Am. J. Phys.***68**(8): 723-727

Harzer, P. 1908. Die Sterne und der Raum. *Jahresbericht der Deutschen Mathematiker-Vereinigung***17**: 237-267

Hoefer, C. 1994. Einstein's struggle for a Machian gravitation theory. *Stud. His. Phil. Sci.***25**(3): 287-335

Hubble, E. 1929. A relation between distance and radial velocity among extra-galactic nebulae. *PNAS.***15**: 168-173

Janssen, M. 2005. Of pots and holes: Einstein's bumpy road to general relativity. *Ann. Physik***14** (Suppl S1) 58-85

Janssen M. and J. Renn 2007. Untying the knot: how Einstein found his way back to field equations discarded in the Zurich notebook. In *The Genesis of General Relativity Vol 2.* M. Janssen et *al.* (eds). *Boston Studies in the Philosophy of Science***250**: 839-925. Springer, Berlin.

Kerzberg, P. 1989. The Einstein-de Sitter controversy of 1916-1917 and the rise of relativistic cosmology. *In Einstein and the History of General Relativity, Einstein Studies Vol 1. D. Howard and J. Stachel (eds) Birkhäuser pp 325-366*

Kragh, H. 2007. *Conceptions of Cosmos*. Oxford University Press, Oxford

Kragh, H. 2012a. Geometry and astronomy: pre-Einstein speculations of non-Euclidean space. ArXiv preprint 1205.4909

Kragh, H. 2012b. Is space flat? Nineteenth-century astronomy and non-Euclidean geometry. *J Ast. Hist. Herit.***15**(3): 149-158

Kragh, H. 2018. Hubble Law or Hubble-Lemaître Law? The IAU Resolution. ArXiv preprint 1809.02557

Lemaître, G. 1927. Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annal. Soc. Sci. Brux.* Série A. **47:** 49-59

Lemaître, G. 1958. Recontres avec Einstein. *Rev. Quest. Sci.***129:** 129-132

Lobachevsky, N.I. 2010. *Pangeometry.* In *Source Book of Mathematics.* D.E. Smith (ed*).* McGraw Hill. Dover (1984): 360-371. European Mathematical Society

Milne, E. A. 1933. World-Structure and the Expansion of the Universe. *Zeit. Astrophys.***6**: 1-95

Milne, E. A. 1935. *Relativity, Gravitation and World Structure*. Clarendon Press, Oxford

Neumann, C. 1896. *Allgemeine Untersuchungen über das Newtonsche Prinzip der Fernwirkungen*. Teubner, Leipzig.

Newcomb, S. 1906. The structure of the universe. In *Sidelights on Astronomy and Kindred Fields of Popular Science*. Harper & Bros, New York

North, J.D. 1965. *The Measure of the Universe: A History of Modern Cosmology*. Oxford University Press, Oxford. Republished by Dover in 1990.

Norton J. 1984. How Einstein found his field equations: 1912-1915. *Hist. Stud. Phys. Sci.***14:** 253-316

Norton J. 1999. The cosmological woes of Newtonian gravitation theory. In *The Expanding Worlds of General Relativity, Einstein Studies Vol. 7.* H. Goenner et al. (eds) Birkhäuser pp 271-322

Nussbaumer, H. and L. Bieri. 2009. *Discovering the Expanding Universe*, Cambridge University Press, Cambridge

Nussbaumer, H. 2014a. Einstein's aborted model of a steady-state universe. To appear in the volume "In memoriam Hilmar W. Duerbeck" *Acta Historica Astronomiae.* Ed. W. Dick, R. Schielicke and C. Sterken. http://arxiv.org/abs/1402.4099

Nussbaumer, H. 2014b. Einstein's conversion from his static to an expanding universe. *Eur. Phys. J (H)***39**(1): 37-62

O'Raifeartaigh, C. 2013. The contribution of V.M. Slipher to the discovery of the expanding universe. In *'Origins of the Expanding Universe: 1912-1932'* Ed. M. Way and D. Hunter, Astron. Soc. Pacific. Conf. Series **471:** 49-61

O'Raifeartaigh, C. and B. McCann. 2014. Einstein's cosmic model of 1931 revisited; an analysis and translation of a forgotten model of the universe. *Eur. Phys. J (H)***39**(1): 63-85

O'Raifeartaigh, C., B. McCann, W. Nahm and S. Mitton. 2014. Einstein's steady-state theory: an abandoned model of the cosmos. *Eur. Phys. J (H)***39**(3):353-367

O'Raifeartaigh, C., O'Keeffe M., Nahm W. and S. Mitton. 2017. Einstein's 1917 static model of the universe: a centennial review. *Eur. Phys. J (H)***42**(3): 431-474

O'Raifeartaigh, C. and S. Mitton. 2018. Interrogating the legend of Einstein's "biggest blunder". *Physics in Perspective.***20**(4): 318–341

Peirce, C.S. 1891. The architecture of theories. *The Monist***1**: 161-176

Realdi, M. and G. Peruzzi 2009. Einstein, de Sitter and the beginning of relativistic cosmology in 1917. *Gen. Rel. Grav.***41** (2): 225-247

Schrödinger, E. 1918. Über ein Lösungssystem der allgemein kovarianten Gravitationsgleichungen. *Phys. Zeit.***19:** 20-22

Schulmann R., Kox A.J., Janssen M. and J. Illy. 1998. *The Collected Papers of Albert Einstein Volume 8. The Berlin Years: Correspondence 1914-1918* (Transl. A.M. Hentschel). Princeton University Press, Princeton.

Schwarzschild, K. 1900. Über das zulässige Krümmungsmass des Raumes. *Vierteljahrschrift Ast. Gesell.***35**: 337-347. Or 'On the permissible curvature of space'. *Class. Quant. Grav*. **15**: 2539-2544 (1998).

Schwarzschild, K. 1916. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinischen Theorie. *Sitz. König. Preuss. Akad. Wiss*. 189-196. Or 'On the gravitational field of a mass point according to Einstein's theory' *Gen. Rel. Grav*. **35** (5): 951-959 (2003).

Seeliger, H. von 1895. Über das Newtonische Gravitationsgesetz. *Astron. Nach*. **137**: 129-136.

Seeliger, H. von 1896. Über das Newtonische Gravitationsgesetz *Sitz. König. Bayer. Akad. Wiss*. **126**: 373-400

Selety, F. 1922. Beitrage zum kosmologischen Problem. *Ann. Physik***68**: 281-334

Smeenk, Christopher. 2014. Einstein's role in the creation of relativistic cosmology. In *The Cambridge Companion to Einstein*. M. Janssen and C. Lehner (eds). Cambridge University Press, Cambridge pp 228-269

Thomson, W. 1901. On ether and gravitational matter through infinite space. *Phil. Mag*. **2**: 160-177

Zöllner, J.C.F. 1872. *Über die Natur der Cometen*. Engelmann, Leipzig

Stability in Cosmology, from Einstein to Inflation

C. D. McCoy

Abstract I inquire into the role of stability in cosmology by investigating two episodes from the recent history of cosmology: (1) Einstein's static universe and Eddington's demonstration of its instability and (2) the flatness problem of the hot big bang model and its alleged solution by inflationary theory. These episodes illustrate differing reactions to instability in cosmological models, both positive ones and negative ones. To provide some context to these reactions, I situate them in relation to perspectives on stability from dynamical systems theory and its epistemology. This reveals, among other things, that an insistence on stability is an extreme position in light of the broad spectrum of physical systems exhibiting degrees of both stability and fragility, one which has perhaps a pragmatic rationale but not any deeper one.

1 Introduction

The meeting in Bern where this paper was presented commemorated the first applications of Einstein's general theory of relativity, in particular, marking, among other things, the 100th anniversary of the publication of Einstein's famous paper "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie" (Cosmological Considerations in the General Theory of Relativity) (Einstein, 1917), the founding paper of the field of relativistic cosmology. In this paper, Einstein proposed an unchanging, temporally infinite, and spatially finite relativistic model of the large-scale universe, the so-called Einstein static universe. Other cosmological models of general relativity soon followed, for example, by de Sitter (1917a,b,c), Friedman (1922, 1924), and Lemaître (1927). Despite ever-increasing observational evidence that the universe is not static and is in fact expanding (as these latter models allowed), Einstein firmly maintained his belief in his static model for many years,

C. D. McCoy (✉)
Underwood International College, Yonsei University, Seoul, South Korea
e-mail: casey.mccoy@yonsei.ac.kr

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_4

finally publicly disavowing it only in a 1931 paper (Einstein, 1931). The decisive point for Einstein was not any observational evidence however; rather, he seems to have been ultimately convinced by Eddington of the unsuitability of his model due to its *instability* (Nussbaumer, 2014; O'Raifeartaigh and McCann, 2014).

So began a minor, albeit sometimes significant thread through the history of modern cosmology, tying together episodes involving the theme of stability. It is this theme which is my topic. While there are, admittedly, many greater themes in the history of cosmology, stability has nevertheless played appreciable and occasionally important roles in the development of cosmology, as I aim to show.[1] How attitudes to stability have changed over the years is of special interest, particularly because of the great advances in the study of nonlinear dynamics in the twentieth century (such as the development of chaos theory), advances which crucially depend on the presence of instability in a model or system. In pursuing my aims, I choose to examine two important episodes from the history of cosmology: the Einstein–Eddington episode just mentioned (Sect. 2) and the more recent 'flatness problem' of the big bang model (Sect. 3), which helped usher in the contemporary paradigm of inflationary cosmology in the 1980s.

As said, Einstein himself was ultimately convinced to abandon his cosmological model on stability grounds. Other physicists, such as Eddington and Lemaître, rather than simply abandoning it, sought to make use of this instability as a means for effecting a transition to an expanding universe, through some manner of physical perturbation. Although the static universe idea maintained some currency in the following decades, it was eventually dropped altogether by the great majority of cosmologists in favor of the past-finite expanding universe models of Friedman and Lemaître, in particular in the guise of the big bang model, developed especially by Gamow and his collaborators in the 1940s.

Later, however, the big bang model itself was brought into question due to its alleged 'fine-tuning problems', which led to the widespread adoption of inflationary theory in the 1980s on the grounds (at least initially) that it solved these problems. One of these problems is the flatness problem. The flatness problem begins with observations which have increasingly suggested that the spatial geometry of the universe is nearly flat (or Euclidean). The spatially flat big bang models are dynamically unstable however: any slight deviation from flatness results in an increasingly divergent curvature. Yet this instability was not given by cosmologists as the sole reason to seek an alternative model. Rather, as a consequence of this instability, they inferred from the observed high degree of flatness an extraordinary degree of fine-tuning in the initial conditions of the universe. It is this fine-tuning that was taken by cosmologists as the reason to reject the simple big bang model and implement a dynamical means (inflation) for insuring, among other things, the stability of spatial flatness.

[1] In this vein, a referee and the editors of this volume encouraged the mention of the classic paper by Lifshitz (1946) on gravitational instabilities.

Even this quick sketch evinces an interesting variety of attitudes toward stability held by the physicists involved. Nevertheless, when one scrutinizes their arguments, it is not so apparent why they took the attitudes that they did. It is not clear, for example, what they took the physical significance of instability (or lack thereof) to be, how to characterize the physical source of the perturbations they imagine, which variations matter, how to ground the 'improbability' of finely tuned states, and so on. Greater historical and individual context are surely needed to answer these questions fully. Yet some insight can be gained by unearthing a once commonly accepted conceptual paradigm: that stability is a necessary feature of physical systems (and models based on them). The ubiquity of this paradigm can perhaps help account for some reactions to the instability of the Einstein static universe, such as Einstein's.

As time went on, however, this paradigm came to be seen by many as an outmoded piece of dogma, thanks to the aforementioned advances in dynamical systems theory. It became ever more evident that one should not, indeed *cannot*, expect dynamical models to be stable, except in very simple cases. Models exhibiting aspects of what is sometimes called *fragility*—some degree of instability—have found important applications in physics and beyond, a particularly visible example, again, being those models studied in chaos theory. Many ideas and much mathematics concerning such models were developed before the twentieth century, yet the widespread realization of their physical significance and ubiquity has grown out of research in nonlinear dynamical systems only since the 1960s. In the final section of the paper (Sect. 4), I draw especially on ideas and arguments presented in Tavakol (1991) concerning these developments in order to assess and reflect on the theme of stability in cosmology, particularly the two episodes which are my focus in this paper.

2 The Einstein Static Universe

The first episode is the story of Einstein's static universe model. Besides introducing this model, the first relativistic spacetime model of the universe, (Einstein, 1917) is also famous for his modification of the field equations to include the cosmological constant.[2] He was moved to make this modification in light of what he took to be a certain paradox in cosmological models involving gravitation, both Newtonian and relativistic. This paradox is not so relevant to the discussion here, nor are most of Einstein's motivations for introducing his cosmological model and the cosmological constant.[3] Thus, rather than following Einstein's fairly idiosyncratic presentation of his model, I will specify the Einstein static universe in modern geometrical terms, for this will help make the structure of the model and its instability more manifest.

[2] For commentaries on this paper, see Smeenk (2014) and O'Raifeartaigh et al. (2017).

[3] See, e.g., Norton (1992), Earman (2001), and Kragh (2007) for discussions of this background.

In these terms, one understands a relativistic spacetime as a geometrical model of the general theory of relativity (Hawking and Ellis, 1973; Wald, 1984; Malament, 2012). It is a differentiable manifold $\mathcal{M}$ equipped with a Lorentzian metric field g. The matter content of the spacetime is specified by the stress energy tensor field T and the cosmological constant is denoted by the scalar field Λ. The metric field g and stress energy field T associated with $\mathcal{M}$ satisfy the Einstein field equations with cosmological constant (EFE-Λ):

$$R - \frac{1}{2}\mathcal{R}g - \Lambda g = 8\pi T, \tag{1}$$

where R is the Ricci tensor and $\mathcal{R}$ is the Ricci scalar, both obtainable from g. The Einstein static universe can then be specified as the non-vacuum ($T \neq 0$) relativistic spacetime, which satisfies the following conditions: (1) it is *spatially homogeneous and isotropic* and (2) it is *static*.

A spacetime that is spatially homogeneous and isotropic possesses a congruence of timelike curves such that it is possible to foliate spacetime by a one-parameter family of spacelike hypersurfaces orthogonal to the timelike curves. Given this foliation, spatial homogeneity and isotropy jointly require that there exists some timelike vector field such that for each spatial hypersurface Σ_λ, where λ denotes the temporal parameter along the timelike curves, the geometrical characteristics of all points in each such hypersurface are the same.[4] With these symmetries, the matter content of the spacetime can be represented as a perfect fluid with energy density ρ and pressure p. The EFE-Λ will then reduce to two coupled equations (the Friedman equations):

$$3H^2 + 3\kappa = 8\pi\rho + \Lambda; \tag{2}$$

$$3\dot{H} + 3H^2 + 4\pi(\rho + 3p) = \Lambda, \tag{3}$$

where H is the Hubble parameter, specifying the expansion of space, and κ is the curvature of space.[5] Only Λ is assumed to remain constant in time (note that derivatives with respect to the temporal parameter are denoted by overdots). The spacetimes thus obtained are generally called the Friedman(–Lemaître)–Robertson–Walker (FRW or FLRW) models.

Among the FRW models, the Einstein static universe is distinguished by the second condition: staticity. A static FRW spacetime is one where $H = 0$; that is, it

---

[4] A more precise characterization of these conditions are given in Malament (2012, §2.11). See also McCabe (2004) and McCoy (2016).

[5] The Hubble parameter H is related to the expansion scalar θ (from the Raychaudhuri equation) by $\theta = 3H$. The spatial curvature κ is related to the Ricci scalar $\mathcal{R}_\Sigma$ of the spatial hypersurfaces in FRW spacetimes by $\mathcal{R}_\Sigma = -6\kappa$. Cf. Malament (2012, §2.11). Note that my κ is the negative of Malament's $\mathcal{K}$, and he uses normal and script fonts to differentiate the spacetime and spatial Ricci tensors and scalars, respectively, where I introduce subscripts for the spatial ones.

neither expands nor contracts. This simplifies the EFE-Λ equations further to

$$3\kappa = 8\pi\rho + \Lambda; \tag{4}$$

$$4\pi(\rho + 3p) = \Lambda. \tag{5}$$

Einstein's paper considers only the case where matter is pressureless 'dust'. In this special case ($p = 0$), the universe is static when $\Lambda = 4\pi\rho = \kappa$.[6] It also follows that space is positively curved in the static universe, since $\kappa = 4\pi\rho > 0$, and that Λ is a positive cosmological constant. For if $\kappa = 0$, then $\rho = \Lambda = 0$— we have Minkowski spacetime, a spacetime without matter— and if $\kappa < 0$, then $\rho < 0$, violating the weak energy condition. Thus, the only non-vacuum spacetime that satisfies the Einstein static universe conditions is the positively curved one. Since the only possible spatial topology of such a spacetime is that of the sphere, it follows that the Einstein static universe is spatially finite, a circumstance which Einstein found particularly favorable.[7]

In the more general case, where $p \neq 0$, we may repeat the previous argument, obtaining essentially the same results. If $\kappa = 0$, then $\rho = -p$ and the 'matter' content of the universe acts as an inverse cosmological constant, which offsets the 'actual' cosmological constant Λ. In other words, we have just found Minkowski spacetime again. If $\kappa < 0$, then $\rho < 0$ or $\rho + p < 0$, in either case violating the weak energy condition. Thus, only in the case that $\kappa > 0$ do we have a physically reasonable non-vacuum spacetime, and, again, it is one that is spatially finite.

In the same year as Einstein proposed the static universe model, de Sitter proposed an alternative model of the universe (de Sitter, 1917a,b,c). For the sake of comparison, it is worth describing his model briefly. The de Sitter universe is a vacuum spacetime (at large scales, anyway), which expands at a constant rate due to the presence of, in effect, a (positive) cosmological constant. Assuming spatial homogeneity and isotropy again, for the sake of comparison to FRW models, the EFE-Λ in this case reduce to

$$3H^2 + 3\kappa = \Lambda; \tag{6}$$

$$3\dot{H} + 3H^2 = \Lambda. \tag{7}$$

If we choose a foliation of de Sitter spacetime where the spatial hypersurfaces are flat ($\kappa = 0$), we see that $\dot{H} = 0$ and

$$3H^2 = \Lambda; \tag{8}$$

[6]Cf. Einstein (1923, 187). See also Malament (2012, 194).

[7]Thus, $\mathcal{M} = S^3 \times \mathbf{R}$, which is why the Einstein static universe is sometimes called the 'cylindrical universe'.

that is, spatial expansion (given our arbitrary choice of what space is in the de Sitter universe) is constant and proportional to the (positive) cosmological constant. As a consequence, geodesics diverge from one another exponentially in time. Also, like the Einstein static universe, time in the de Sitter universe is past and future infinite; that is, the de Sitter universe neither begins nor ends.

Much of the debate in the 1920s focused on these two cosmological models (Kragh, 1996). Despite the lack of matter in de Sitter's universe, it made a stronger connection to available cosmological observations, especially by furnishing a possible explanation for Slipher's observations of redshifting in the spectral lines of galaxies (Slipher, 1912, 1915, 1917, 1921). Einstein was unaware of the latest observational results in astronomy and instead relied on his intuition about the nature of the universe, holding 'philosophically' that it was static and spatially finite. Uncertainty and confusion about coordinate choices and the geometrical structure of the models, as well as about the actual physical features of the universe accessible through observation, stymied understanding for many years. Perhaps for these reasons, it was not until end of the decade that the instability of the Einstein static universe was noticed and communicated by Eddington (1930).[8]

Eddington considers only a small scalar perturbation of the density ρ in a universe consisting of dust ($p = 0$). Combining the Friedman equations by eliminating the curvature term, we have

$$\dot{H} = \Lambda - 4\pi\rho. \tag{9}$$

As noted above, the Einstein static universe is the universe where $4\pi\rho = \Lambda$. Clearly a "slight disturbance" which causes ρ to increase or decrease will lead the universe to contract or expand, respectively, without returning to and oscillating around the static solution. As Eddington concludes, "evidently Einstein's universe is unstable" (Eddington, 1930, 670).

We might also consider the case where the pressure is positive, in which case we have

$$\dot{H} = \Lambda - 4\pi(\rho + p). \tag{10}$$

A slight disturbance in the density or the pressure would, however, again clearly lead to a changing Hubble parameter and a demonstration of the model's instability.

One might easily question Eddington's demonstration on a couple of points. First, why should the relevant equations of motion be those pertaining to the FRW class of spacetimes and not all relativistic spacetimes? Eddington considers only a single kind of perturbation, one which is consistent with such universes: a homogeneous and isotropic spacelike perturbation of the energy density. This is

[8] Eddington does, however, acknowledge Lemaître's investigation into the Einstein and de Sitter universes (Lemaître, 1927), with respect to which he says, "it is at once apparent from his formulæ that the Einstein world is unstable" (Eddington, 1930, 668).

enough to show that the Einstein static universe is unstable if we suppose that a model is unstable when it is unstable with respect to any perturbation.

Still, one might wonder whether such a perturbation is physically significant. Eddington and other cosmologists thought so for a time, adopting a vision of the universe as having been in a past-infinite state described by Einstein's model, but which at some point transitions via a perturbation to an expanding epoch (and eventually becoming approximated by de Sitter's model after much expansion) (Robertson, 1933). What physical process could lead to a perturbation of this kind in the Einstein static universe?

In typical physical systems, a model that is dynamically unstable is liable to be pushed out of the unstable state by physical perturbations from the environment. Indeed, one might even say that it is *improbable* for a system to persist in such a state for any appreciable amount of time due to the presence of a highly perturbing environment. But this idea cuts no ice in cosmology, as there is obviously no external environment from which perturbations impinge on the universe: the perturbation could not come from 'outside the universe'. As Eddington himself says, such a perturbation would be "supernatural".

Nevertheless, he claims that "the initial small disturbance can happen without supernatural interference" (Eddington, 1930, 670). The proposal he initially moots is that the gravitational collapse of "uniformly diffused nebula" into galaxies would lead to just such a perturbation: "the actual mass may not alter but the equivalent mass to be used in applying the equations for a strictly uniform distribution must be slightly altered" (Eddington, 1930, 670). However, insofar as no new mass is created in gravitational collapse, it is difficult to see why the 'equivalent mass' used in the Friedman equations would change during this process. If the volume of space does not change, and the amount of matter in space does not change, how could the energy density change? This could occur under some conversion process of matter into radiation, which, as Eddington notes, would not change ρ but would change p. If such a conversion were to occur, though, it ought to lead to an *increase* in ρ or p, which would lead to spatial *contraction*.[9] As the aforementioned redshift observations by Slipher already suggested, along with Hubble's estimates of the distances to galaxies (Hubble, 1929), space should be expanding in the model, not contracting. And a realistic physical mechanism which could result in this, the Einstein static universe falling out of its static 'equilibrium' into an expanding universe, remained quite elusive.

Nevertheless, Eddington's argument that the static universe was unstable convinced Einstein to abandon his model, as the latter acknowledged in Einstein (1931). That paper is perhaps better known as the paper in which he abandons the "unsatisfactory" cosmological constant, instead favoring the spherical, expanding

[9] This point was investigated soon after Eddington's discovery, especially by McVittie and McCrea. See McCrea and McVittie (1930, 1931), McVittie (1931), Dingle (1933), Tolman (1934), Sen (1935a,b). Lemaître proposed an alternative mechanism for the departure from equilibrium, which he called a "stagnation" (Lemaître, 1931).

FRW spacetime discovered first by Friedman (1922).[10] In the paper, Einstein states two principal reasons for abandoning his static universe: (1) that it was unstable, as Eddington had convinced him previously in conversation, and (2) due to the observational results of Hubble.[11] Nevertheless, although he acknowledges the observational results indicating the expansion of the universe, he insists that the instability of the model was *reason enough* not to ascribe physical significance to the static universe: "schon aus diesem Grunde bin ich nicht mehr geneigt, meiner damaligen Lösung eine physikalische Bedeutung zuzuschreiben, schon abgesehen von Hubbels Beobachtungsresultaten" (Einstein, 1931, 236).[12]

One might prescind from the evident conceptual significance of instability to the participants in this episode, noting that the outcome was nothing more than the adoption of models that were more empirically adequate than the Einstein static universe (and the de Sitter universe)—simply a paradigmatic example of empirical progress guiding theoretical progress. This is clearly not how the participants reasoned however. Moreover, it is of some interest that their reactions to the instability discovered by Eddington differed. Einstein, for example, saw instability as undermining the physical significance of a model, whereas Eddington, Lemaître, and others saw it as an opportunity to introduce a mechanism for explaining the generation of complex structure in the universe.

Unfortunately, none of them discusses his particular views on the physical significance of stability, so it is difficult to reconstruct any concrete argument which they would have made. One possible line of argument, which might nevertheless be gleaned from various scattered remarks in these papers, begins with the obvious: the universe is not perfectly uniform, despite it being assumed so in the Einstein static universe and the other cosmological models of the time. If we represent deviations from uniformity as perturbations in our models and suppose that the actual universe should be a 'perturbation away' from our favored idealized model, then the perturbed model which represents the actual universe should not deviate significantly from the static model (or other idealized model) over time. Otherwise, the static model is a poor approximation of the actual universe and inapt for prediction, explanation, etc. The instability of the Einstein static universe suggests just this: any slight deviation from the Einstein static universe results in significant qualitative differences between the perturbed model and the unperturbed model.

This objection, however, is not by itself sufficient to reject the static model outright. So long as the static model is a sufficiently good approximation to the actual universe *for some cosmological epoch*, it can be used to describe and make

[10]One year later, Einstein and de Sitter (1932) argued instead for the flat, expanding FRW spacetime, on the grounds that there is no direct observational evidence for nonzero spatial curvature. This model, the Einstein–de Sitter universe, became the standard model of cosmology for much of the twentieth century and will reappear in the following section.

[11]See O'Raifeartaigh and McCann (2014) for a translation and discussion of this paper. See also Nussbaumer (2014) for further analysis of Einstein's reasons.

[12]"On these grounds alone, I am no longer inclined to ascribe a physical meaning to my former solution, quite apart from Hubbel's observations" (O'Raifeartaigh and McCann, 2014, 83).

predictions about the universe during that epoch. Of course, as it happens, the static model fits very poorly with observations *throughout* the history of the cosmos, and this surely is reason enough to reject it outright. Yet if it had turned out that observations fit it well enough for the present epoch, what reason could instability provide to abandon it, at least as an approximation?

It appears that rejecting the Einstein static universe merely on the grounds that it is unstable can only derive from a (dogmatic) insistence on stability in physical models. In contrast to this negative attitude, Eddington and other cosmologists endeavored to hold on to the static universe, at least for a time, despite a lack of any observational evidence for it (and presumably because they entertained certain 'philosophical' reasons), recognizing that its instability could be used to effect a 'phase transition' in the large-scale universe. Thus, we can see in this episode both a degree of "positivization" of instabilities (Schmidt, 2011, 223) in cosmology and what appears to be a complete rejection of them in line with what has been called the 'dogma of stability' (Abraham and Marsden, 1978, xxii), that is, an insistence that only stable models have physical significance.

3 The Flatness Problem

My second episode is the story of the flatness problem. It is one of the fine-tuning problems that led to the widespread adoption of inflationary theory by theoretical cosmologists in the 1980s, due to inflationary theory's alleged solution of them. As mentioned above, the flatness problem begins with observations suggesting that the universe's spatial curvature is approximately flat (when the universe is modeled with a FRW spacetime). The flatness problem arises in part because the Friedman equations can be used to demonstrate the dynamic instability of flat curvature under small perturbations, much in the same way that Eddington showed that the Einstein static universe was unstable. The problem is not just the presence of this instability however. The crucial problem, according to proponents of inflation, is that the initial conditions of the universe had to be extremely fine-tuned, due to this instability, for the universe to be anywhere near as flat as observations suggested. It is this fine-tuning that they reject, not the instability per se. Inflationary theory purports to solve this fine-tuning problem with a short stage of exponential expansion in the very early universe (inflation), which reverses the dynamical stability of FRW universes, thereby making flat FRW spacetimes dynamically stable under perturbations (at least during inflation).

The big bang model of the universe is based on the expanding FRW spacetimes, introduced in the previous section. These may have positive, negative, or flat spatial curvature. This curvature is determined by the density and pressure of spacetime's contents. It has been an important empirical matter in the later twentieth century to determine what the (large-scale) curvature of space. Although observations have long suggested that the flat model (the 'Einstein–de Sitter universe' of Einstein and de Sitter (1932), as it is sometimes called) is roughly accurate, it cannot be excluded that space has a positive or negative curvature (nor could it, since observations do not have infinite accuracy). Indeed, for aesthetic reasons, many physicists have in the past preferred a spatially finite, (slightly) positively curved model (the 'Einstein–Friedman universe' of Friedman (1922) and Einstein (1931)). Nevertheless, it is this observed approximate flatness that is the empirical fact on which the flatness problem is based.

To see how the problem arises, let us revisit the basic assumptions which lead to the FRW spacetimes. First, we suppose that there is a congruence of timelike curves, such that spacetime can be foliated by a one-parameter family of spacelike hypersurfaces ('Weyl's principle'). Then we assume that there exists a foliation where the spacelike hypersurfaces are homogeneous and isotropic ('the cosmological principle'). The spacetimes satisfying these conditions are the FRW spacetimes. They obey the Friedman equations:

$$3H^2 + 3\frac{k}{a^2} = 8\pi\rho; \tag{11}$$

$$3\dot{H} + 3H^2 + 4\pi(\rho + 3p) = 0, \tag{12}$$

where I have replaced the curvature κ used previously by the expression k/a^2. The parameter k specifies the sign of the curvature: +1 for positively curved, -1 for negatively curved, and 0 for flat. The scale factor a parameterizes the expansion and curvature of space; it is related to the Hubble parameter H by $H = \dot{a}/a$. In this model, the 'big bang' itself occurs at $a = 0$ and the present at $a = 1$ (by convention).[13]

A flat FRW spacetime has $k = 0$. Only a specific value of the energy density ρ will result in a universe with exactly flat spatial curvature. This is the *critical density* ρ_{cr}. It is obtained from the first Friedman equation by setting k to zero:

$$\rho_{cr} = \frac{3}{8\pi}H^2. \tag{13}$$

The present Hubble parameter H_0 and the present density ρ_0 may be determined from observations, thus allowing the comparison of ρ_0 and ρ_{cr}. These have been determined to be extremely close in recent years, although their approximate equality has been accepted for many decades. Thus, if our modeling assumptions are correct, we appear to live in a (very nearly) spatially flat universe.

It is straightforward to demonstrate the dynamical instability that features in the flatness problem. One way is to modify the first Friedman equation slightly, dividing it by the critical density ρ_{cr} and defining a new parameter, the density parameter $\Omega = \rho/\rho_{cr}$. Then one has

[13] Of course, the singularity is not a point in this spacetime, so $a = 0$ is strictly speaking not a valid parameter value of the scale factor (although any $a > 0$ is).

$$1 - \Omega = -\frac{k}{(aH)^2}. \tag{14}$$

As we would like to see what happens when there are small departures from flatness and we do not care whether the departures are positive or negative, we may for convenience take the absolute value of both sides and ignore the $k = 0$ case. We then have the following equation:

$$|1 - \Omega| = \frac{1}{(aH)^2}. \tag{15}$$

So long as the universe is expanding, the right-hand side is always increasing in time. Its time derivative is

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{(aH)^2} = -\frac{2\ddot{a}}{\dot{a}^3}. \tag{16}$$

If the universe only has normal matter in it, then $\ddot{a}$ is always negative (normal matter gravitates and hence decelerates expansion). Therefore, we may conclude that $|1 - \Omega|$ increases in time under small perturbations from flatness, and increasingly so.[14]

The instability of flat FRW spacetimes allows a fine-tuning argument to be made based on it.[15] Our universe is presently observed to be expanding with a nearly flat spatial geometry. If it is exactly flat, then its initial conditions in the very early universe were such that it had exactly the critical density. If, however, it had ever so slightly different initial conditions, such that it had slightly less or slightly more than the critical density, then it would be nowhere near spatially flat today: it would be highly curved, in most cases to a degree that would not permit our existence. In other words, only initial conditions in a very narrow range would result in the presently observed universe. The big bang model, in short, requires that our universe's initial conditions be highly fine-tuned. One can do various calculations to get a sense of the degree of fine-tuning; Baumann (2009, 23), for example, calculates the fine-tuning to be one part in 10^{55} for initial conditions placed at the GUT scale.

There are various reactions one might have to the fact of this fine-tuning. One is, "so what?" If we trust our models and observations, then it is simply a logical consequence that the universe had to have had such-and-such initial conditions, within a range suggested by the uncertainty in observations. What does it matter that they could not have been much different? After all, presumably any physical system

[14]For positively curved FRW spacetimes, this is true only up to a point. These universes reach a maximum curvature, after which they contract into a 'big crunch'.

[15]Fine-tuning arguments take different forms and invoke varying considerations, as the general notion of fine-tuning arises in other contexts besides relativistic cosmology, for example, high energy physics (Williams, 2015) and the notorious 'fine-tuning for life' problem debated by scientists, philosophers, and the religious. See Friederich (2018) for an introduction to these latter kinds. Dicke apparently was responsible for popularizing the flatness problem among physicists. See, e.g., Dicke and Peebles (1979).

requires *some* particular initial conditions. Such an attitude appears to be partly behind the analysis of Earman and Mosterín (1999, 19–20), and it is presumably the reaction that many other philosophers of science, especially strongly empiricist ones, would take.

This is not the attitude that theoretical cosmologists take. They say that this fine-tuning indicates that the initial conditions required by the hot big bang model are special and therefore problematic. Unfortunately, they are not so clear about what exactly makes fine-tuned initial conditions problematic (McCoy, 2015). Cosmologists usually interpret this specialness in terms of likelihoods or probabilities. However, there are serious problems, both conceptual and technical, with interpreting fine-tuning problems in cosmology in this way (Callender, 2004; McCoy, 2018a). If cosmologists are instead merely reporting on their subjective degrees of belief, then it is unclear why anyone else should take their pronouncements very seriously, that is, as a matter of physical significance. There should, in short, be some objective justification of likelihood attributions for them to have physical meaning. Since likelihoods or probabilities are not part of the theory of general relativity, though, it is quite difficult to see from where they might come.[16]

In keeping with the negative reactions to instability in the previous episode, one might maintain that the specialness of fine-tuned initial conditions in the flatness problem is owed simply to the instability of the dynamics for the condition of flatness. At least in this case it would be justified to say that flatness is special, for under the FRW dynamics it has a special property: instability. Moreover, if this were the right way to interpret fine-tuning in this case, then inflationary theory does solve the problem, by reversing the instability of flatness (McCoy, 2015). The difficulty, however, is the same as that raised in the previous episode: namely, sustaining the claim that this instability is problematic. Here as elsewhere, it is not clear what the argument would be for insisting on stability.

Cosmologists in the 1980s and afterward generally do not take the instability of flatness to be a problem as such however. Rather, the instability of flatness coupled with the observed approximate flatness of the universe entails that the big bang universe had to have special initial conditions, and it is this latter fact which they take to be problematic. Therefore, the reaction to instability is somewhat different in this episode compared to the previous one. It is just a single factor contributing to the fact which actually raises concerns about the physical significance of the model. That said, it is clear that in both cases instability is taken to have important consequences for theorizing in cosmology. True, what consequences those are differ in the episodes and among those involved, especially depending on whether instability is seen by them as a vice or a virtue.

[16] Interestingly, several physicists have made likelihood-based arguments that there is actually no flatness problem (Gibbons et al., 1987; Hawking and Page, 1988; Coule, 1995; Gibbons and Turok, 2008; Carroll and Tam, 2010). Their arguments, however, suffer from problems more or less as serious as those who claim that there is a flatness problem. For criticism, see Schiffrin and Wald (2012) and McCoy (2017).

4 Stability and Fragility

To provide greater insight into the physical significance of instability, in this section I consider it in a more general context, namely in dynamical systems theory. The precise notion of stability in dynamical systems theory is usually attributed to Andronov and Pontryagin, in their article "Grubye sistemy" (Coarse systems) (Andronov and Pontryagin, 1937), although this was preceded by important work by Poincaré and Birkhoff, among others.[17] Around the same time as this latter pioneering early work was being done on dynamical systems theory, however, the epistemic significance of stability was being emphasized by, among others, Duhem (1962, Part 2, Ch. 3) and Poincaré (1952, I.IV.II)—that is, even in the face of a growing appreciation of the mathematical significance of instability in dynamical systems.

While the history is worthy of investigation, for convenience I will make use of the narrative given by Tavakol (1991) to situate some of the main philosophical considerations. Tavakol argues that scientists generally assume, usually implicitly, that both real systems and mathematical models of those systems are structurally stable. In other words, they adopt the aforementioned 'dogma of stability'. He characterizes this assumption as follows (Tavakol, 1991, 148):

(a) Real systems are structurally stable in the sense that they do not change their qualitative behavior under small general perturbations.
(b) For mathematical models to be viable as models of real systems, they similarly need to be structurally stable.

So, for example, in the context of cosmology, this amounts to assuming that the universe does not change its qualitative behavior under small general perturbations and mandating that our mathematical models respect this supposed fact. Since the Einstein static universe certainly does change its qualitative behavior under even a small scalar perturbation, changing from a static to a non-static universe, it violates the assumption. Similarly, since the Einstein–de Sitter universe changes from statically flat spatial curvature to dynamically changing spatial curvature, it too is not in accord with the assumption.

Tavakol offers two important reasons for adopting these assumptions, which he calls the 'stability framework'. One is that the framework addresses the problem of relating idealized models to real systems. The other is that the framework addresses the inaccuracy of observations.

First, idealization. There are a number of reasons that scientists use idealized models. Some are practical. For example, one must invoke simplifying assumptions to make solving actual problems tractable (particularly if the dynamics is nonlinear). Some reasons have a more epistemic character. As described in section one, in the case where we use a simplified model to stand in for a real system, we would like

[17] For historical accounts of the development of dynamical systems, see, e.g., Aubin and Dahan-Dalmedico (2002) and Holmes (2010).

to know that any differences between the model and the system will not undermine the model's predictions, explanations, etc. For example, our universe is clearly not exactly spatially homogeneous and isotropic, but we model it as if it were; if the universe were structurally stable, then its small departures from homogeneity and isotropy would not change its qualitative behavior of being (nearly) uniform.

Second, inaccuracy. As Tavakol says, the stability framework "facilitate[s] the task of interpreting observational and experimental data, by providing a theoretical framework… within which such data can be analysed" (Tavakol, 1991, 148). It does so because real observations always involve some degree of inaccuracy due to error. The stability framework addresses this issue by insuring that small inaccuracies in measurement will not undermine the application of models whose parameters depend on these measurement results. So long as the model accurately representing the real world is nearby ('a perturbation away'), then we can expect that predictions based on using the 'inaccurate' model will be approximately correct.

Nevertheless, assumption (a) is not an insubstantial assumption, and it has plausibly been adopted largely for pragmatic reasons in physics. As mentioned previously, that this is so has become especially clear in the past century thanks to the recognition of the phenomenon of chaos in nonlinear dynamical systems, the applicability of which to real systems demonstrates that real systems are not necessarily structurally stable. Chaotic systems, characterized by a sensitive dependence on initial conditions and a mixing dynamics, entail a failure of deterministic predictability in chaotic models representing them (Batterman, 1993; Holt and Holt, 1993; Leiber, 1997; Werndl, 2009). Chaotic systems, then, are examples of systems that violate the assumptions of the stability framework. In such cases, the stability framework cannot be used to account for idealizations and measurement inaccuracy. Indeed, chaotic systems present a significant challenge for the interpretation of empirical data and accounting for the confirmation of chaotic models (Rueger and Sharp, 1996; Koperski, 1998; Batitsky and Domotor, 2007). One must appeal to alternative methods and justifications to secure the epistemic goods that were previously licensed by adopting the stability framework.

If we must give up the stability framework, then Tavakol claims that there are two alternatives: (1) hold onto dynamical determinism but adopt what he calls the 'fragility framework' or (2) maintain stability at the expense of determinism by introducing stochasticity. Each of these raises interesting philosophical challenges. On the one hand, adopting the fragility framework amounts to accepting that real systems may be structurally unstable, as a consequence of which our models must allow for the corresponding kinds of instabilities.[18] As just noted, this requires a different approach to understanding our modeling and testing practices. On the other hand, Tavakol's example of maintaining stability by introducing stochasticity

[18] It is important to recognize that stability is not an all or nothing affair: one expects a spectrum of possibilities, in principle, between highly stable and highly fragile. Indeed, this in mind, Kamminga and Tavakol (1993) suggest interpreting ceteris paribus clauses in laws of nature as identifying, roughly, the absence of relevant variability, which is to say the absence of the kind of perturbations which would lead to changes in qualitative behavior.

involves a redefinition of stability to allow for it. A kind of stochastic stability is certainly a natural alternative to conventional stability, as Poincaré (1952, I.IV.II) and others pointed out long ago. If this stochasticity is to be objective, however, then there is a philosophical challenge to taking up this strategy, namely of grounding the objectivity of these deterministic probabilities (Lyon, 2011; Myrvold, 2012; McCoy, 2018b).

We are forced into this dilemma in cosmology not just because many of the cosmological models that we know are unstable—but because we should expect instability in general in cosmology: it is a general feature of nonlinear dynamical systems, of which the systems described by general relativity are a kind.[19] So, either we accept the fragility framework in cosmological modeling (and its attendant philosophical challenges) or we adopt a statistical approach in cosmology (and its attendant philosophical challenges).[20] If this dynamical systems perspective is right, then the rejection of a cosmological model, like the Einstein static universe, solely on the basis of its instability under certain perturbations is surely epistemically unjustified. It is, at best, a pragmatic decision which one might make in the hopes of discovering a stable model, or, at least, a model stable under perturbations deemed relevant for assessing whether the idealizations of the model are acceptable.

5 Conclusion

Determinism, probability, and causality are concepts that have long played important roles in scientific, physical, and cosmological thinking; they have also been central to philosophical discussions. The aim of this paper has been to draw some attention to the significant role that stability considerations have played in cosmology too. My focus has been on two well-known cases, the Einstein static universe and the flatness problem, which exhibit interesting aspects of stability in this historical thread. It would be a mistake, of course, to drive any strong, general conclusions on the basis of just these cases. Instead, I hope to have shown enough to suggest the conceptual significance of stability in cosmology, some of the roles it can play, and how attitudes toward it have varied and changed.

In the first section, I introduced the Einstein static universe and the de Sitter universe and derived some important consequences, such as the spherical geometry of space and the condition for stability of the former. I then showed how Eddington demonstrated the instability of the Einstein static universe to scalar perturbations of the energy density ρ and pressure p. I related some of the historical consequences of Eddington's demonstration, including Einstein's disavowal of his model and the search by many cosmologists for a physical mechanism which would introduce

[19] This does not necessarily mean that we should expect cosmology to be chaotic, although that is a conceptual possibility that has been investigated. See, e.g., Coleman and Pietronero (1992).

[20] See Tavakol and Ellis (1988), Tavakol and Ellis (1990), Coley and Tavakol (1992), and Lidsey and Tavakol (1993) for Tavakol's own application of his arguments to cosmology, including several examples.

a perturbation leading to an expanding universe, as observations suggested our universe was doing. Although empirical evidence provided the strongest argument for rejecting the Einstein static universe, it seems that Eddington's argument played an important role—certainly it did for Einstein.

In the second section, I explained the flatness problem, which arises in the context of the hot big bang model. Observations suggest that our universe is approximately spatially flat, but the spatially flat FRW spacetime is unstable under perturbations. This led some cosmologists to press a fine-tuning argument against the model, concluding that the special initial conditions required to account for flatness were problematic in a way that required a theoretical modification of the big bang model. The proposed solution to the flatness problem was inflation, a stage of accelerated expansion in the very early universe. Although instability was not singled out as the problem of the flat FRW spacetime, it is only because of the particular kind of instability that the initial conditions are as special as they are, whether that specialness is understood in terms of probability or otherwise. The inflationary solution reverses the dynamics, so to speak, in such a way that flatness becomes stable and the initial conditions become less special.

In the third section, I contextualized these two episodes by discussing stability in light of advances in dynamical systems theory, especially in nonlinear dynamics and chaos theory. I made use of Tavakol's alternative frameworks, the stability framework and fragility framework, discussing the epistemic issues associated with each one. Tavakol draws particular attention on the consequences of stability and fragility for idealization and inaccuracy. One important conclusion that can be drawn from this general context is that there are only pragmatic reasons to work within the stability framework, for it has become clear that physical systems, or, at least, our best models of them, do in fact exhibit instabilities of various kinds.

Reflecting back on the cosmological episodes, we can see that an outright rejection of the Einstein static universe is epistemically unjustified (although it perhaps can be motivated on pragmatic grounds). The response of Eddington, Lemaître, and others, to make this instability an opportunity, appears to be better motivated; in light of the subsequent fine-tuning arguments against the big bang model, however, we can easily imagine their proposal being subject to fine-tuning arguments of various kinds—for example, in the posited equilibrium state, in the required perturbation to obtain an expanding universe of the appropriate kind, and so on. This might make fine-tuning arguments appear spurious, and indeed this is a(n especially philosophical) reaction to the later flatness problem. Although characterizing what is problematic about special conditions in terms of instability or improbability cannot be sustained, as I have argued elsewhere (McCoy, 2018a), I also believe the empirical successes of inflationary cosmology should not be overlooked. Indeed, the theory did solve conceptual problems which did lead to progress (McCoy, 2019). It is just not yet clear precisely *how* to characterize these problems conceptually. One suggestion of how, which I believe is worth pursuing, comes out of the discussion of stability here. That is, that the instability of flat FRW spacetimes points to the lack of explanatory robustness of the idealizations that lead to it.

References

Abraham, Ralph and Jerrold Marsden. 1978. *Foundations of Mechanics, 2nd edn*. Redwood City: Addison Wesley.

Andronov, Aleksandr and Lev Pontryagin. 1937. Grubye sistemy. *Doklady Akademii Nauk SSSR* 14:247–251.

Aubin, David and Amy Dahan-Dalmedico. 2002. Writing the history of dynamical systems and chaos: *Longue Durée* and revolution, disciplines and cultures. *Historia Mathematica* 29(3):273–339.

Batitsky, Vidam and Zoltan Domotor. 2007. When good theories make bad predictions. *Synthese* 157(1):79–103.

Batterman, Robert. 1993. Defining chaos. *Philosophy of Science* 60(1):43–66.

Baumann, Daniel. 2009. TASI lectures on inflation. Available via arxiv.org. http://arxiv.org/abs/0907.5424. Cited 31 July 2018.

Callender, Craig. 2004. Measures, explanations and the past: Should 'special' initial conditions be explained? *The British Journal for the Philosophy of Science* 55:195–217.

Carroll, Sean and Haywood Tam. 2010. Unitary evolution and cosmological fine-tuning. Available via arxiv.org. http://arxiv.org/abs/1007.1417. Cited 31 July 2018.

Coleman, Paul and Luciano Pietronero. 1992. The fractal structure of the universe. *Physics Reports* 213:311–389.

Coley, Alan and Reza Tavakol. 1992. Fragility in cosmology. *General Relativity and Gravitation* 24(8):835–855.

Coule, David. 1995. Canonical measure and the flatness of a FRW universe. *Classical and Quantum Gravity* 12:455–469.

de Sitter, Willem. 1917a. On Einstein's theory of gravitation and its astronomical consequences, III. *Monthly Notices of the Royal Astronomical Society* 78:3–28.

de Sitter, Willem. 1917b. Over de kromming der ruimte. *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling Verslagen van de Gewone Vergaderingen* 26:222–236.

de Sitter, Willem. 1917c. Over de relativiteit der traagheid: Beschouwingen naar aanleiding van Einsteins laatste hypothese. *Koninklijke Akademie van Wetenschappen te Amsterdam Wis- en Natuurkundige Afdeeling Verslagen van de Gewone Vergaderingen* 25:1268–1276.

Dicke, Robert and Jim Peebles. 1979. The big bang cosmology—enigmas and nostrums. In *General Relativity: An Einstein Centenary Survey*, eds. Stephen Hawking and Werner Israel, 504–517, Cambridge: Cambridge University Press.

Dingle, Herbert. 1933. On isotropic models of the universe, with special reference to the stability of the homogeneous and static states. *Monthly Notices of the Royal Astronomical Society* 94:134–158.

Duhem, Pierre. 1962. *The Aim and Structure of Physical Theory*. New York: Atheneum.

Earman, John. 2001. Lambda: The constant that refuses to die. *Archive for History of Exact Sciences* 55:189–220.

Earman, John and Jesus Mosterín. 1999. A critical look at inflationary cosmology. *Philosophy of Science* 66:1–49.

Eddington, Arthur. 1930. On the instability of Einstein's spherical world. *Monthly Notices of the Royal Astronomical Society* 90:668–678.

Einstein, Albert. 1917. Kosmologische betrachtungen zur allgemeinen relativitätstheorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*: 142–152

Einstein, Albert. 1923. Cosmological considerations on the general theory of relativity. In *The Principle of Relativity*, eds. Wilfrid Perrett and George Jeffrey, 177–188, London: Methuen and Co.

Einstein, Albert. 1931. Zum kosmologische problem der allgemeinen relativitätstheorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*: 235–237.

Einstein, Albert and Willem de Sitter. 1932. On the relation between the expansion and the mean density of the universe. *Proceedings of the National Academy of Sciences* 18:213–214.

Friederich, Simon. 2018. Fine-tuning. In *The Stanford Encyclopedia of Philosophy, Spring 2018*, ed. Edward Zalta. https://plato.stanford.edu/archives/spr2018/entries/fine-tuning/. Cited 31 July 2018.

Friedman, Aleksandr. 1922. Über die krümmung des raumes. *Zeitschrift für Physik* 10(1):377–386.

Friedman, Aleksandr. 1924. Über die möglichkeit einer welt mit konstanter negativer krümmung des raumes. *Zeitschrift für Physik* 21(1):326–332.

Gibbons, Gary and Neil Turok. 2008. Measure problem in cosmology. *Physical Review D* 77(063516).

Gibbons, Gary, Stephen Hawking, and John Stewart. 1987. A natural measure on the set of all universes. *Nuclear Physics B* 281:736–751.

Hawking, Stephen and George Ellis. 1973. *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Hawking, Stewart and Don Page. 1988. How probable is inflation? *Nuclear Physics B* 298:789–809.

Holmes, Philip. 2010. A short history of dynamical systems theory: 1885–2007. In *History of Mathematics*, eds. Vagn Hansen and Jeremy Gray, 115–138, Singapore: Eolss.

Holt, Lynn and Glynn Holt. 1993. Regularity in nonlinear dynamical systems. *The British Journal for the Philosophy of Science* 44(4):711–727.

Hubble E (1929) A relation between distance and radial velocity among the extra-galactic nebulae. Proceedings of the National Academy of Sciences 15:168–173.

Kamminga, Harmke and Reza Tavakol. 1993. How untidy is god's mind? A note on the dynamical implications of Nancy Cartwright's metaphysics. *The British Journal for the Philosophy of Science* 44(3):549–553.

Koperski, Jeffrey. 1998. Models, confirmation, and chaos. *Philosophy of Science* 65(4):624–648.

Kragh, Helge. 1996. *Cosmology and Controversy*. Princeton: Princeton University Press.

Kragh, Helge. 2007. *Conceptions of the Cosmos*. Oxford: Oxford University Press.

Leiber, Theodor. 1997. On the actual impact of deterministic chaos. *Synthese* 113(3):357–379.

Lemaître, Georges. 1927. Un univers homogène de mass constant et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles Série A* 47:49–59.

Lemaître, Georges. 1931. The expanding universe. *Monthly Notices of the Royal Astronomical Society* 91:490–501.

Lidsey, James and Reza Tavakol. 1993. On the correspondence between theory and observations in inflationary cosmology. *Physics Letters B* 309(1–2):23–27.

Lifshitz, Evgenii. 1946. On the gravitational stability of the expanding universe. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 16: 587–602.

Lyon, Aiden. 2011. Deterministic probability: neither chance nor credence. *Synthese* 182:413–432.

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

McCabe, Gordon. 2004. The structure and interpretation of cosmology: Part I—general relativistic cosmology. *Studies in History and Philosophy of Modern Physics* 35:549–595.

McCoy, Casey. 2015. Does inflation solve the hot big bang model's fine-tuning problems? *Studies in History and Philosophy of Modern Physics* 51:23–36.

McCoy, Casey. 2016. Philosophical implications of inflationary cosmology. PhD thesis.

McCoy, Casey. 2017. Can typicality arguments dissolve cosmology's flatness problem? *Philosophy of Science* 84(5):1239–1252.

McCoy, Casey. 2018a. The implementation, interpretation, and justification of likelihoods in cosmology. *Studies in History and Philosophy of Modern Physics* 62:19–35.

McCoy, Casey. 2018b. No chances in a deterministic world. Available via philsci-archive.pitt.edu. http://philsci-archive.pitt.edu/13166/. Cited 31 July 2018.

McCoy, Casey. 2019. Epistemic justification and methodological luck in inflationary cosmology. *The British Journal for the Philosophy of Science* 70(4):1003–1028.

McCrea, William and George McVittie. 1930. On the contraction of the universe. *Monthly Notices of the Royal Astronomical Society* 91:128–133.
McCrea, William and George McVittie. 1931. The expanding universe. *Monthly Notices of the Royal Astronomical Society* 92:7–12.
McVittie, George. 1931. The problem of n bodies and the expansion of the universe. *Monthly Notices of the Royal Astronomical Society* 91:274–283.
Myrvold, Wayne. 2012. Deterministic laws and epistemic chances. In *Probability in Physics*, eds. Yemima Ben-Menahem and Meir Hemmo, 73–85, Berlin: Springer.
Norton, John. 1992. A paradox in newtonian gravitation theory. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992:412–420.
Nussbaumer, Harry. 2014. Einstein's conversion from his static to an expanding universe. *European Physical Journal H* 39:37–62.
O'Raifeartaigh, Cormac and Brendan McCann. 2014. Einstein's cosmic model of 1931 revisited: an analysis and translation of a forgotten model of the universe. *European Physical Journal H* 39(1):63–85.
O'Raifeartaigh, Cormac, Michael O'Keeffe, Werner Nahm, and Simon Mitton. 2017. Einstein's 1917 static model of the universe: a centennial review. *European Physical Journal H* 42(3):431–474.
Poincaré, Henri. 1952. *Science and Method*. New York: Dover.
Robertson, Howard. 1933. Relativistic cosmology. *Reviews of Modern Physics* 5(1):62–90.
Rueger, Alexander and David Sharp. 1996. Simple theories of a messy world: Truth and explanatory power in nonlinear dynamics. *The British Journal for the Philosophy of Science* 47(1):93–112.
Schiffrin, Joshua and Robert Wald. 2012. Measure and probability in cosmology. *Physical Review D* 86(023521).
Schmidt, Jan. 2011. Challenged by instability and complexity. In *Philosophy of Complex Systems*, ed. Cliff Hooker, 223–254, Oxford: Elsevier.
Sen, Nikhil. 1935a. On a minimum property of the Friedmann space. *Zeitschrift für Astrophysik* 9:315–318.
Sen, Nikhil. 1935b. On the stability of cosmological models with nonvanishing pressure. *Zeitschrift für Astrophysik* 10:291–296.
Slipher, Vesto. 1912. The radial velocity of the andromeda nebula. *Lowell Observatory Bulletin* 2(8):56–57.
Slipher, Vesto. 1915. Spectrographic observations of nebulae. *Popular Astronomy* 23:21–24.
Slipher, Vesto. 1917. Nebulæ. *Proceedings of the American Philosophical Society* 56(5):403–409.
Slipher, Vesto. 1921. Two nebulae with unparalleled velocities. *Popular Astronomy* 29:128.
Smeenk, Chris. 2014. Einstein's role in the creation of relativistic cosmology. In: The Cambridge Companion to Einstein, eds. Michel Janssen and Christoph Lehner, 228–269, Cambridge: Cambridge University Press.
Tavakol, Reza. 1991. Fragility and deterministic modelling in the exact sciences. *The British Journal for the Philosophy of Science* 42:147–156.
Tavakol, Reza and George Ellis. 1988. On the question of cosmological modelling. *Physics Letters A* 130:217–224.
Tavakol, Reza and George Ellis. 1990. Stability of cosmological models. *Physics Letters A* 143(1–2):8–12.
Tolman, Richard. 1934. *Relativity, Thermodynamics and Cosmology*. Oxford: Clarendon Press.
Wald, Robert. 1984. *General Relativity*. Chicago: University of Chicago Press.
Werndl, Charlotte. 2009. What are the new implications of chaos for unpredictability? *The British Journal for the Philosophy of Science* 60(1):195–220.
Williams, Porter. 2015. Naturalness, the autonomy of scales, and the 125gev Higgs. *Studies in History and Philosophy of Modern Physics* 51:82–96.

The Einstein–Rosen Bridge and the Einstein–Podolsky–Rosen Argument: Singularities and Separability

Galina Weinstein

Abstract In 1935 Einstein pursued two main paths separately: the Einstein–Rosen (ER) bridge theory and the Einstein–Podolsky–Rosen (EPR) argument. In this paper, I deal with the static two-particle problem in general relativity and relationship of this problem with the two projects on which Einstein worked in parallel. I discuss two questions: What was the possible role played by the static two-body problem in the rise and fall of the ER bridge theory? What was the possible role played by the static two-body problem in Einstein's formulations of the EPR argument? Finally, I also briefly discuss a possible link between Einstein's work on EPR and the ER bridge.

1 Introduction

In October 1933, Einstein moved to Princeton. He was occupied with unified field theory and criticized the new quantum mechanics. From the end of 1933 to the spring of 1935, Einstein used the methods of general relativity to account for atomic particles and electric phenomena in a particular way: together with his young assistant Nathan Rosen he invented a solution which later became known as the ER bridge to represent an elementary material particle by using only the field equations of general relativity while at the same time avoiding the occurrence of any singularities. Unification of field theory with elementary particles of matter was achieved through bridges in spacetime (van Dongen, 2010, 131). The ER bridge paper was published in the spring of 1935 (Einstein and Rosen, 1935).

Einstein's work on unified field theory went hand-in-hand with his refusal to accept that quantum mechanics was a complete theory. Accordingly, Einstein endeavored to adapt quantum mechanics to the requirements of the general theory of relativity. During that spring of 1935, Einstein also published a paper with Rosen

G. Weinstein (✉)
Department of Philosophy, University of Haifa, Haifa, Israel

Israel and Interdisciplinary Center (IDC), Herzliya, Israel
e-mail: galina.weinstein@post.idc.ac.il

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_5

and Boris Podolsky in which they propounded the well-known EPR argument on the incompleteness of quantum mechanics (Einstein et al., 1935).

Let us fast forward many years to find physicists Juan Martín Maldacena and Leonard Susskind suggest that the ER bridge is a special kind of the EPR correlation (the ER = EPR conjecture). According to this conjecture, these two concepts are, in fact, related by more than the common publication date of 1935. If two particles are connected by entanglement, then they are effectively joined by an ER bridge and vice versa, an ER bridge is equivalent to entanglement (Maldacena and Susskind, 2013).

From the historian's point of view, we can imagine Einstein in 1935, sitting in his office having discussions with his assistant Rosen on unified field theory, general relativity, and problems in quantum mechanics. Podolsky, a member in the school of mathematics in 1934–1935, joined the conversations. Indeed, according to Leopold Infeld's autobiography, Einstein probably stood up in his office, his assistants standing near him, he took a piece of chalk, went to the blackboard and started to deliver a lecture on the basic ideas underlying his unified field theory (Infeld, 1941, 254–259). As discussed in Sect. 4.2, a letter by Einstein to Michele Besso from 1936 suggests that Podolsky also worked with Einstein on the ER bridge problem.

Section 2 provides a brief historical introduction to the static two-body problem in general relativity. In Sect. 3, I discuss Einstein's response to these solutions and present Ludwig Silberstein's solution for the static two-body problem. In 1933, Einstein started to receive letters from Silberstein, in which he presented to him his solution.

Einstein responded to Silberstein's solution in his ER bridge paper. In Sect. 4, I present the theory of the ER bridges and, in Sect. 4.2, I discuss the relationship between the two-body problem in general relativity and the ER bridge. In Sect. 5.2 I try to answer the following question: What was the possible role played by the static two-body problem in the development of the ER bridge theory?

Section 5 presents the EPR argument, the EPR paper, and Einstein's own formulations of the EPR argument. I discuss the relationship between the static two-particle problem in general relativity and Einstein's formulations of the EPR argument. In Sect. 5.2 I try to answer the following question: What was the possible role played by the static two-body problem in Einstein's formulations of the EPR argument?

2 Weyl, Levi–Civita and Bach: Static Two-Body Problem in General Relativity

Hermann Weyl (1917, 137–142) and Tulio Levi-Civita (1919) both invented a method to solve Einstein's field equations for the case of a static, axially symmetric field.

Using what Weyl called "canonical" cylindrical coordinates r, θ, z, the line element was written in the following form (Levi-Civita, 1919, 10):

$$ds^2 = e^{2\psi} dt^2 - e^{-2\psi}[e^{2\gamma}(dr^2 + dz^2) + r^2 d\theta^2], \tag{1}$$

where the functions, ψ and γ, only depend on canonical coordinates r and z. Insertion into Einstein's vacuum field equations:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0, \tag{2}$$

with $R_{\mu\nu}$ representing the Ricci tensor and R the Ricci scalar, implies that ψ is a solution of the following *linear* equation (Levi-Civita, 1919, p. 9; Weyl, 1917, p. 137):

$$\nabla^2 \psi = \frac{\partial^2 \psi}{\partial z^2} + \frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r}\frac{\partial \psi}{\partial r} = 0, \tag{3}$$

which represents Laplace's equation for an axisymmetric function ψ. Once ψ is obtained, we can then determine γ, which is a solution of two non-linear equations:

$$\frac{\partial \gamma}{\partial r} - r\left[\left(\frac{\partial \psi}{\partial r}\right)^2 - \left(\frac{\partial \psi}{\partial z}\right)^2\right] = 0, \qquad \frac{\partial \gamma}{\partial z} - 2r\frac{\partial \psi}{\partial r}\frac{\partial \psi}{\partial z} = 0. \tag{4}$$

Just as in Newtonian mechanics, one can assume, heuristically, that ψ is a Newtonian potential generated by a certain source with axial symmetry expressed in cylindrical coordinates in a flat Euclidean space. However, the non-linearity of Eqs. (4) renders the physical interpretation of the solution in terms of real space highly non-trivial.

Rudolf (Förster) Bach used the above method to find the Schwarzschild solution for the field of a single particle of mass m. Bach took ψ in the form of a Newtonian potential of a rod, with total mass m and length $2l$, located along the z-axis (Bach and Weyl, 1922, 134):

$$\psi_{12} = \frac{m}{2} \ln \frac{r_1 + r_2 - 2l}{r_1 + r_2 + 2l}, \tag{5}$$

where r_1 and r_2 are the distances from the ends of the rod to the point in the field.

Integration of Eqs. (4) gives

$$\gamma_{12} = \frac{m}{2} \ln \frac{(r_1 + r_2)^2 - 4l^2}{4r_1 r_2}. \tag{6}$$

Inserting ψ_{12} and γ_{12} into Levi-Civita's (1) yields the Schwarzschild metric in axially symmetric coordinates.

Any superposition of two solutions ψ of the linear Laplace equation (3) is also a solution of (3). But this is not the case for the corresponding γ, which is a solution of Eqs. (4). Bach then used the canonical line element (1) to study the superposition of two Schwarzschild solutions in axially symmetric coordinates. He obtained, for this case, the Newtonian potential $\psi = \psi_{12}+\psi_{34}$ produced by two rods located between points z_1 and z_2 and between points z_3 and z_4, respectively, with $z_1 > z_2 > z_3 > z_4$ located on the z-axis. ψ_{12} is then given by (5) and ψ_{34} has a similar form to ψ_{12} for the second rod located between z_3 and z_4.

The non-linear equations (4) yield the solution: $\gamma = \gamma_{12} + \gamma_{34} + \gamma_{23}$. Again γ_{12} is given by (6) and γ_{34} has a similar form to γ_{12} corresponding to the second rod.

Calculating γ_{23}, Bach noted that the static solution for the two-body problem implies the presence of a singularity Γ on the line connecting the two bodies. It arises from the third term, γ_{23}, representing the interaction between the two bodies. This fact violates the regularity of the solution and the bodies cannot be in equilibrium under the influence of gravitational forces alone (Bach and Weyl, 1922, 138, 141).

In 1922, Weyl published remarks, “The Static Two Body Problem,” as an addendum to Bach’s paper, in which he set himself the task to solve this problem (Bach and Weyl, 1922, 142–144). In 1919, Weyl had already assumed that masses were held at rest by stresses that compensate for the gravitational forces of the masses and he now calculated the components of the stresses that hold the masses at rest from Levi-Civita’s equations (Weyl, 1919, 186–188).

For bodies at rest (static solutions), Weyl established that the components of the stress-energy tensor density $\mathfrak{T}_\tau^\nu = \sqrt{g}T_\tau^\nu$ represent the radial axial stresses (that compensate for the gravitational forces) and the azimuthal stresses (which hold the bodies fixed). The condition for the radial axial stresses is

$$\mathfrak{T}_1^1 + \mathfrak{T}_2^2 = 0. \tag{7}$$

Further, one has

$$\mathfrak{T}_2^1 - \mathfrak{T}_1^2 = 0, \tag{8}$$

and the mass density ρ and the energy density ρ' are given by $\mathfrak{T}_3^3 = r\rho'$, $\mathfrak{T}_0^0 = r(\rho + \rho')$.

The radial axial stresses can be calculated from the stress-energy tensor $T_{\mu\nu}$ in Einstein’s field equations:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = T_{\mu\nu}. \tag{9}$$

In 1922, Weyl took into consideration $\rho' = 0$, $\mathfrak{T}_0^0 = r\rho$. Then the function ψ satisfies a Poisson linear equation in canonical coordinates (r, z):

$$\nabla^2 \psi = \frac{1}{2}\rho, \tag{10}$$

and ψ is uniquely split into two parts, $\psi_1 + \psi_2$.

Between these two bodies the mass density is $\rho = 0$ but $\mathfrak{T}_\tau^\nu \neq 0$.

Using (4), (7), and (8), Weyl avoided the singularity Γ along the z-axis between the two bodies by radial axial stresses that compensate for the gravitational forces.

3 Einstein on the Two-Body Problem

3.1 Einstein and Grommer: The Impossible Is Impossible

In 1927, in a paper with his assistant Grommer, Einstein used the Weyl–Levi-Civita method to find the spacetime of a mass point. Einstein and Grommer took ψ in the form of the Newtonian potential of a mass point m lying on the z-axis at $z = 0$ (Einstein and Grommer, 1927, p. 5):

$$\psi_1 = -\frac{m_1}{R_1}, \tag{11}$$

where $R_1 = \sqrt{r^2 + z^2}$. Integration of Eqs. (4) gives:

$$\gamma_{11} = -\frac{m_1^2 r^2}{2R_1^4}. \tag{12}$$

This solution is singular at $R_1 = 0$. Einstein, however, easily remedied this problem by locating the singularity at the origin of the cylindrical axis, where one finds the mass point. It follows that the singularity can be regarded as describing an elementary material particle and therefore one has $\gamma = 0$ on the positive and negative z-axis.

Now suppose that in addition to the field arising from the singularity (12) the particle is at rest in an external gravitational field having axial symmetry. We put

$$\psi = \psi_1 + \bar{\psi}, \tag{13}$$

where ψ_1 is given by (11).

$\bar{\psi}(r, z)$ is assumed regular around $r = z = 0$, it represents the potential of the external gravitational field, and it has the following form:

$$\bar{\psi} = \alpha_0 + \alpha_1 z + G, \tag{14}$$

where G includes terms of second and higher order in r and z.

Einstein and Grommer now argued that Eq. (13) corresponds to a two-body solution of (3). Equations (4) yield a solution such that if γ vanishes on one side of the singularity (at $z = 0$), it is constant on the other side. We can therefore set $\gamma = 0$ on the negative z-axis but in order for the solution to be regular for all z (except at $z = r = 0$) γ must also vanish along the positive z-axis. This will only be the case if the integral $d\gamma$ over a closed semi-circle around the singularity (at $z = r = 0$) vanishes. The calculation provides the condition that $\alpha_1 = 0$ in order to get $\oint d\gamma = 0$. This condition is fulfilled if the external gravitational field vanishes, $\bar{\psi} = 0$, at $z = r = 0$, the location of the material particle. This represents a situation which is almost a *reductio ad absurdum* of the field equations. Thus, (13) is not regular and requires the presence of a singularity on the positive z-axis and it is assumed that outside a particle at rest in an external field, space (r, z) is not free of singularities.

As Weyl had wrestled with the static two body problem, he felt a need to introduce what in later literature became known as the Weyl strut. He said in his 1922 Addendum: "The physical significance of this interpretation [the Weyl strut] should not be exaggerated. For the solution of the real two-body problem, nothing was gained by the determination of the motion of two gravitating bodies attracted towards each other" (Bach and Weyl, 1922, p. 145). By these words he meant that although Eq. (3) for ψ is linear, the two equations (4) for γ are non-linear. Thus, the static solution for the two-body problem always requires the presence of a singularity between the two bodies.

Einstein and Grommer mentioned in a footnote Weyl's (1917), Weyl's (1919), Levi-Civita's (1919), and Bach and Weyl's (1922) papers but they did not refer explicitly to Weyl's 1922 addendum to Bach's paper (Einstein and Grommer, 1927, p. 5). Einstein thus may not have been aware of the above quote from Weyl's addendum.

3.2 *Silberstein 1933: The Impossible Is Possible*

In 1933, Silberstein used the Weyl–Levi-Civita method to find the spacetime of two mass points. He took ψ in the form of the Newtonian potential of two mass points of masses m_1 and m_2, lying on the z-axis at $z = d_1$ and $z = -d_2$, and derived the equation:

$$\psi = \psi_1 + \psi_2. \tag{15}$$

ψ_1 is given by (11) and ψ_2 is given by

$$\psi_2 = -\frac{m_2}{R_2}, \tag{16}$$

where now $R_1 = \sqrt{r^2 + (z - d_1)^2}$ and $R_2 = \sqrt{r^2 + (z + d_2)^2}$.

Silberstein integrated the field equations (4), took into consideration (12), and obtained the solution:

$$\gamma = \gamma_{11} + \gamma_{22} + \gamma_{12} = -\frac{r^2}{2}\left(\frac{m_1}{R_1^2} + \frac{m_2}{R_2^2}\right)^2 + \gamma_{12}, \tag{17}$$

where $\gamma_{12} = \frac{2m_1m_2}{d}\left(\sqrt{1 - \frac{d^2r^2}{R_1^2R_2^2}} - 1\right)$, and d is the separation between the two masses which lie along the z-axis.

The fact is that in 1924, this "solution for two singularities" on the z-axis was first suggested independently by Jean Chazy (1924, pp. 26–27, 35–36) and Harry Curzon (1924, pp. 477–478). But Silberstein was apparently not acquainted with their work.

The odd thing was that Silberstein came to a directly opposite conclusion to that of Einstein and Grommer: the Chazy–Curzon solution (which Silberstein rediscovered), or Eq. (15), does not require singularities on the line connecting the two static masses (Silberstein, 1936, p. 268).

Before submitting his results as a paper to the *Physical Review*, Silberstein communicated them in 1933 to Einstein. According to Silberstein, Eqs. (15) and (17) are exact solutions of the vacuum field Eqs. (2). Equation (17) has only two singularity points, $R_1 = 0$ and $R_2 = 0$, that are located at the positions of the mass centers of the two material bodies, m_1 and m_2. However, it does not contain any singularities along the line connecting the two bodies. It would be unreasonable to assume, as did Weyl, that the two masses are forced to remain at relative rest by stresses, i.e. by a Weyl strut. For this would mean the existence of a stress-energy tensor according to the field equations (9). But (17) is not a solution of (9), it is an exact solution of the vacuum field Eqs. (2). In trying to find an explanation of (17), Silberstein argued that his solution was inadmissible physically and contradicted experience. But it was perfectly possible to write such an exact solution of Eqs. (2). In light of this extraordinary outcome, Silberstein felt he had no other choice but to tell Einstein that not every solution of (2) would be physically admissible.[1]

At first Einstein told Silberstein that his solution is singular and not valid but then he added: "a really complete [field] theory would exist only if 'matter' could be represented in it by fields and without singularities."[2] Silberstein tried to persuade Einstein that his solution was indeed non-singular outside the location of the two point masses. Overcoming his objections to Silberstein's solution, Einstein

[1] Einstein to L. Silberstein, 23 September 1935 (AEA 21-074), cited in Havas (1993, p. 101).

[2] "Allerdings wäre eine wirklich vollständige Theorie erst dann vorhanden, wenn die „Materie" in ihr feldmässig repräsentiert und *ohne Singularitäten* darstellbar wäre." A. Einstein to L. Silberstein, 17 December 1933 (AEA 21-061), cited in Havas (1993, p. 102).

immediately wrote back, convinced that Silberstein's solution was all right: "So it is true that there exists a static solution with only two pointlike singularities."[3]

4 The Einstein–Rosen (ER) Bridge

4.1 The ER Bridge Paper

Silberstein's treatment of the static two-body problem prompted Einstein's remark, in his ER bridge paper with Rosen in 1935:

> [...] writers [i.e. Einstein and Grommer, GW] have occasionally noted the possibility that material particles might be considered as singularities of the field. This point of view, however, we [Einstein and Rosen] cannot accept at all. For a singularity brings so much arbitrariness into the theory that it actually nullifies its laws. A pretty confirmation of this was imparted in a letter [in 1933] to one of the authors [Einstein] by L. Silberstein. As is well known, Levi-Civita [(1919)] and Weyl [(1917)] have given a general method for finding axially symmetric static solutions of the gravitational equations. By this method one can readily obtain a solution [Silberstein's two-body solution (15)] which, except for two point singularities lying on the axis of symmetry, is everywhere regular and is Euclidean at infinity. Hence, if one admitted singularities as representing particles one would have here a case of two particles not accelerated by their gravitational interaction, which would certainly be excluded physically (Einstein and Rosen, 1935, p. 73).

Einstein and Rosen thought they had found a solution to this problem, in the context of the ER bridge. They would exclude singularities from the theory, and at the same time elementary material particles would not have to be represented as singularities of the field. This was simply an ideal situation, killing two birds with one stone: the field equations of general relativity no longer possessed the solution containing the Schwarzschild singularity $r = 2m$, and the assumption of particles represented as singularities of the gravitational field was eliminated.

Einstein and Rosen began by modifying the field equations in order to obtain a theory in which singularities of the field are excluded. They required that the determinant of the metric tensor $g = |g_{\mu\nu}|$ be nonvanishing everywhere. In order for g not to vanish, one might at first try to replace the vacuum field equations (2) by modified vacuum field equations (Einstein and Rosen, 1935, p. 74):

$$g^2 R_{\mu\nu} = 0. \tag{18}$$

Einstein and Rosen multiplied (2) by g^2 and avoided the occurrence of denominators in (18). They succeeded in avoiding singularities of that special kind which is characterized by the vanishing of g (Einstein and Rosen, 1935, 74). Schwarzschild's exact spherical symmetric solution in Johannes Droste's form (Einstein and Rosen,

[3] "Es ist also wahr, dass es eine statische Lösung gibt mit nur zwei punktartigen Singularitäten." A. Einstein to L. Silberstein, 24 December 1933 (AEA 21-063), cited in Havas (1993, p. 104).

1935, p. 75),

$$ds^2 = -\frac{1}{1-\frac{2m}{r}}dr^2 - r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2) + \left(1-\frac{2m}{r}\right)dt^2, \quad (19)$$

is a solution of Eq. (18).

In Eq. (19), $1 - \frac{2m}{r}$ becomes infinite at $r = 2m$, and we have the so-called Schwarzschild singularity (a coordinate singularity).

Einstein and Rosen considered the following coordinate transformation: by introducing in place of $r = 2m$ a new variable according to the equation $u^2 = r - 2m$, they removed the region containing the Schwarzschild singularity $r = 2m$, in which $g = |g_{\mu\nu}|$ vanishes. They inserted $r^2 = u^2 + 2m$ into (19) and obtained for ds^2 the expression (Einstein and Rosen, 1935, p. 75):

$$ds^2 = -4(u^2+2m)du^2 - (u^2+2m)^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2) + \frac{u^2}{u^2+2m}dt^2. \quad (20)$$

The Schwarzschild solution (19) becomes a regular solution, free from singularities for all finite points and for all values of u. When u extends from $-\infty$ to $+\infty$, r extends from $+\infty$ to $r = 2m$ and then back to $+\infty$; and for values of $r < 2m$ there are no corresponding real values of u. The solution (20) is a mathematical representation of physical space by a space of two congruent identical flat sheets corresponding to $u > 0$ (which represents one Schwarzschild solution, or this solution is isometric to this sheet) and $u < 0$ (which represents the second Schwarzschild solution). These two sheets are joined by a bridge at $r = 2m$ or $u = 0$.

However, we have avoided the Schwarzschild singularity at $r = 2m$ and have arrived at a new singularity $g = 0$. According to Einstein and Rosen, we can turn off the alarm bells because $g = 0$ neither constitutes a new singularity nor prevents the modified field equations (which have no denominators) from being satisfied. Where $g = 0$, $u = 0$ and the two sheets are connected by the bridge: “In the hypersurfaces of contact of the two sheets the determinant of the $g_{\mu\nu}$ vanishes” (Einstein and Rosen, 1935, p. 77).

Einstein and Rosen tried to replace an elementary particle having mass but no charge with a topological structure, a bridge of finite length in four-dimensional spacetime. Luckily, the spatially finite bridge was identified as being the neutron (discovered in 1932 by James Chadwick), and possibly the neutrino (first postulated by Wolfgang Pauli in 1930), which are elementary particles having mass but no electric charge. With this conception Einstein thought he could represent an elementary particle using only the field equations and not as singularities in the field.

If one had started from the Schwarzschild solution with negative mass m one would have then been unable to make the solution regular, free from singularities, because then $u^2 = r + 2m$. Einstein and Rosen concluded that no ER bridge that corresponds to a neutral particle of negative mass is possible. This is in agreement with the requirement that there can be no neutral particle of negative mass.

Einstein and Rosen now tried to replace an elementary charged particle with a spatially finite bridge by modifying (9) in much the same way as they had done with (2), and then they obtained an equation of a similar nature to (18).

The components of Maxwell's electromagnetic field are related to the electromagnetic potential φ and are represented by an antisymmetric tensor:

$$F_{\mu\nu} = \frac{\partial \varphi_\mu}{\partial x_\nu} - \frac{\partial \varphi_\nu}{\partial x_\mu}. \tag{21}$$

From this it follows:

$$\frac{\partial F_{\mu\nu}}{\partial x_\tau} + \frac{\partial F_{\nu\tau}}{\partial x_\mu} + \frac{\partial F_{\tau\mu}}{\partial x_\nu} = 0, \tag{22}$$

and we can write Maxwell's equations in tensorial form.

These equations have the well-known consequence that the Maxwell stress-energy tensor $T_{\mu\nu}$ in *source free regions (vacuum)* is the following:

$$T_{\mu\nu} = \frac{1}{4} g_{\mu\nu} \varphi_{\alpha\beta} \varphi^{\alpha\beta} - \varphi_{\mu\alpha} {\varphi_\nu}^\alpha. \tag{23}$$

Einstein and Rosen inserted (23) into the field equations (9) and got (Einstein and Rosen, 1935, p. 77):

$$g^2 (R_{\mu\nu} - T_{\mu\nu}) = 0. \tag{24}$$

However, in order to obtain a bridge that would represent a charged particle they were forced to take into consideration the negative of (23). In a footnote they wrote that if $T_{\mu\nu}$ was positive, the solution would involve $+e^2$ instead of $-e^2$. However, it would then not be possible to obtain an ER bridge solution.

They added $-e$ to the Schwarzschild line element (Einstein and Rosen, 1935, p. 77):

$$ds^2 = -\frac{1}{1 - \frac{2m}{r} - \frac{e^2}{2r^2}} dr^2 - r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2) + \left(1 - \frac{2m}{r} - \frac{e^2}{2r^2}\right) dt^2. \tag{25}$$

The above line element is the Reissner–Nordström metric.

Einstein and Rosen wanted to remove the singularity and they set the mass $m = 0$. Thus, (25) has no Schwarzschild limit ($r = 2m$):

$$ds^2 = -\frac{1}{1 - \frac{e^2}{2r^2}} dr^2 - r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2) + \left(1 - \frac{e^2}{2r^2}\right) dt^2. \tag{26}$$

They replaced r (with $m = 0$) by a new variable u according to the equation:

$$u^2 = r^2 - \frac{e^2}{2}, \tag{27}$$

and this removes the region containing the singularity and transforms it into a bridge at $u = 0$:

$$ds^2 = -du^2 - \left(u^2 + \frac{e^2}{2}\right)(d\vartheta^2 + \sin^2\vartheta d\varphi^2) + \frac{2u^2}{2u^2 + e^2}dt^2. \tag{28}$$

Equation (28) is free from singularities for all finite points in the space of two sheets and the charged particle is again represented by a bridge between the two sheets. But desperate Einstein and Rosen had no other choice but to conclude that the bridge represents an elementary electric particle with electric charge but without mass and with negative energy density. We see that in addition to the neutron, which has mass but no electric charge, there appeared to exist according to the ER bridge theory a third type of elementary material particle, a "ghost atom"—as it was called in the press (Laurence, 1935)—an electric particle which has a unit of electric charge but no mass.

Einstein's unified field theory was supposed to account for the existence of two elementary particles: the electron and the proton (Sauer, 2014, p. 214). Evidently, a new difficulty came into the ER bridge theory. Einstein and Rosen obtained two types of bridges connecting the two identical flat sheets. Each type of particle, having either mass or electric charge, is represented by a different type of bridge connecting the two flat sheets. Unlike the "ghost atom," the electron, proton, and positron (anti-electron having the same mass and opposite charge to an electron, predicted by Paul Dirac in 1931) all have both mass and electric charge. "One is therefore led, according to this theory, to consider the electron or proton as a two-bridge problem." The electron and proton would each be represented by two bridges between the two congruent identical flat sheets. Further, "one might expect that processes in which several elementary particles take part correspond to regular solutions of the field equations with several bridges between the two equivalent sheets corresponding to the physical space [...] For the present one cannot even know whether regular solutions with more than one bridge exist at all" (Einstein and Rosen, 1935, p. 77); and thus, Einstein and Rosen failed to account for the electron–proton duo.

After the ER bridge paper, Einstein therefore returned to his previous idea that electrons and protons might be considered as singularities of the field, for reasons mainly connected with failing to find a multi-bridge solution.

4.2 Two-Body Problem and the ER Bridge

4.2.1 The Einstein–Silberstein Controversy

In November 1935, Silberstein was encouraged by finding that Einstein mentioned his name in the ER bridge paper that he sent a paper to the *Physical Review* and opened it by saying:

> The object of this paper is to derive a solution [...] corresponding to two mass centers A, B, a field, that is, which has singularities at A and B only, and not (as in R. Bach's and H. Weyl's physically trivial solution) along the straight segment joining these two points.
>
> I may mention that I have constructed such a solution (a stationary one) in December, 1933 and have then communicated it to Einstein pointing out, rather empathically, that this is a case of a perfectly rigorous solution of the field equations and yet utterly inadmissible physically, so that one cannot henceforth treat 'material particles' as singularities of the field. This has, in fact, induced Einstein to attempt, in collaboration with N. Rosen, a new theory of matter (Silberstein, 1936, pp. 268–269).

Silberstein alluded to his correspondence with Einstein of which we have already caught a glimpse. The foregoing sentence represents Einstein's interpretation of Silberstein's solution. In 1933, Einstein had responded to Silberstein's solution by asserting that a really complete theory would exist only if 'matter' could be represented by fields and without singularities (see Sect. 3.2).

In December 1935, Einstein then wrote to Silberstein and told him that he objected to his solution for the two-body problem, claiming it had an additional singularity on the line connecting the two bodies. But Silberstein became antagonistic to Einstein's general relativity, criticized it, and entered into a controversial debate with Einstein. The controversy between Einstein and Silberstein is related to the question of whether the two-body solution derived by Silberstein stimulated the bridge theory of Einstein and Rosen.

The debate between Silberstein and Einstein continued with more strident tones when Einstein told Silberstein that by his efforts, he had made it necessary for Einstein and Rosen to correct his errors publicly in a letter they were going to send to *The Physical Review*. In February 1936, Einstein's and Rosen's letter was published, as announced, arguing against Silberstein that his solution "fails to satisfy the regularity conditions, for γ is nonvanishing on the axis [...] between the two mass-points" (Einstein and Rosen, 1936, p. 405). Einstein and Rosen admitted though that "the solution given by Silberstein has singularities outside of the two points. This we did not notice in our recent paper (Einstein and Rosen, 1935), where we referred to this solution, which had been previously communicated to one of us" (Einstein and Rosen, 1936, pp. 404–405).

The controversy between Einstein and Silberstein was ventilated during 1936 in the columns of the press, which was far from being a neutral reporter. Articles discussed whether Silberstein actually influenced Einstein and Rosen. In February 1936, *The Gazette* of Montreal announced:

> Professor Albert Einstein, who promulgated his theory of relativity in 1905, has replied to criticism of the theory by Dr. Ludwik Silberstein of the University of Toronto in the current issue of the *Physical Review* [(Einstein and Rosen, 1936)].

The quotation then goes on; Einstein is asked to:

> State his attitude on reports in some newspapers that he was working on a new theory of matter as a result of the 'defects' in his original theory brought to his attention by Dr. Silberstein. Dr. Einstein, in collaboration with Dr. N.Rosen of the Institute of Advanced Studies in Princeton, N.J., published his paper outlining his new theory of matter, linking the theory of relativity and the quantum theory, in the *Physical Review* issue of July 4, 1935. Dr. Einstein's statement makes it clear that Dr. Silberstein's mathematical efforts had nothing to do with the formulating by Dr. Einstein and Dr. Rosen of their new theory of matter, upon which, it has been known, Dr. Einstein has been working for many years.

Silberstein stuck to his solution but finally wrote to "Sweet Mr. Einstein" (AEA 21-088, cited in Havas, 1993, p. 113), and the Silberstein solution was never mentioned again. Afterwards there was a dead silence between Silberstein and Einstein, for a while.

4.2.2 Holes and Bridges

More than ten years later, in 1949, Rosen took up the two-body problem again in a short paper published in *Reviews of Modern Physics*. He added a footnote that indicated: "The above discussion is a generalization of that in the paper by A. Einstein and N. Rosen, *Phys. Rev.* 49, 404 (1936)" (Rosen, 1949, p. 504). Rosen considered a Curzon–Chazy particle at rest in an external gravitational field having axial symmetry, the Einstein–Grommer equation (13). He tried (and failed) to demonstrate that $\gamma = 0$ on the z-axis and came to the same conclusion as that of Einstein and Grommer: for a particle to remain at rest in an external gravitational field, the gravitational force acting on the particle must vanish. Following his mentor's post 1935 ideas, Rosen concluded: "This result might have been expected on the basis of the fact that it has been shown (Einstein, Infeld and Hoffmann, 1938) that the equations of motion of a particle can be derived from the field equation" (Rosen, 1949, pp. 503–504).

Silberstein may well have exaggerated his influence on Einstein but, admittedly, Silberstein did influence Einstein and Rosen, as indeed his name appears in a good generous opening paragraph of their ER bridge paper, where they speak about his static solution and singularities. On February 16, 1936, Einstein explained the ER bridge theory to his good friend Michele Besso. He also confided to him about the many-bridge problem. Einstein began by telling Besso that he hopes to arrive at a really satisfying theory of matter:

> Enclosed I am sending you a short paper, which represents a first step. The neutral and the electric particles appear, so to speak, as a hole in space [Loch im Raume], in such a way that the metric field returns into itself. Space is described as double-sheets. In Schwarzschild's exact spherical symmetric solution, the particle appears in ordinary space as a singularity of the type $1 - \frac{2m}{r}$. Substituting $1 - 2m = u^2$, the field becomes regular in $u - r$-space. When

> u extends from $-\infty$ to $+\infty$, r extends from $+\infty$ to $r = 2m$ and then back to $r = +\infty$. This represents both 'sheets' in Riemann's sense, which are joined by a 'bridge' at $r = 2m$ or $u = 0$. It is similar for the electric charge.
>
> A young colleague (Russian Jew) and I are relentlessly struggling [unablässig schwitze] with the treatment of the many-body problem on that basis (A. Einstein to M. Besso, February 16, 1936, AEA 7-372.1 in (Einstein and Besso, 1972, letter 122)).

Einstein always wished to represent matter in terms of a field theory in which singularities are excluded and the fundamental field equations are the vacuum field equations. His field equations, however, did not have solutions that were free of singularities.

As opposed to a modern distinction between curvature and apparent singularities, for Einstein the most important distinction was rather between singularities that correspond to material particles and those that do not. Einstein, however, could tolerate singularities that represent material particles, but often times he saw how inadequate this type of singularities was (Earman and Eisenstaedt, 1999, pp. 208–220).

Thus, before 1933 and after 1935 Einstein recognized that representing matter by singularities was a simplifying assumption because the field equations were, in fact, sufficient to determine the motion of matter represented as point singularities of the field. The law of motion of the singularities is completely determined by the field equations, without the necessity of an additional law of motion. In 1927, Einstein maintained that this is confirmed by the static two-body solution in axially symmetric coordinates (the Curzon–Chazy metric). In 1938, Einstein also took into consideration "the equations of motion, which we calculate only for the case of two massive particles" (Einstein, Infeld and Hoffmann, 1938, p. 66).

The law of motion of two point singularities is derived from the vacuum field equations (2) alone. However, the Curzon–Chazy metric carries a little problem: between two point singularities (that are located at the positions of the mass centers of the two material bodies) space is not free of singularities (Einstein and Grommer, 1927, p. 5). In 1919 and 1922, Weyl suggested to remove these singularities by radial stresses. But this required to calculate the radial axial stresses from the matter stress-energy tensor $T_{\mu\nu}$ in Einstein's field equations (9), see Sect. 2.

Einstein could not accept this because he was of the opinion that $T_{\mu\nu}$ is a temporary device for representing matter, and the only field equations which follow without ambiguity from the fundamental assumptions of general relativity are the vacuum field equations (2), and it is important to know whether they alone are capable of determining the motion of material particles (Einstein, Infeld and Hoffmann, 1938, p. 65).

But then in autumn 1933 Silberstein showed the opposite for the Curzon–Chazy metric of two bodies. From an ex post facto perspective, Silberstein's solution was unreasonable: a static two-body solution with only two point singularities that does not contain any singularities along the line connecting them. At first Einstein was suspicious; but perhaps it was just wishful thinking that caused him to fall into the trap of a too-good-to-be-true solution.

The neutron (has mass but no charge) and a new massless-charged particle (predicted by the ER bridge theory), each one of them appears as a “hole in space” instead of one point-singularity. The “metric field returns into itself” and forms a hollow bridge. The cavity inside the bridge is the “hole in space.” Space is described as two identical flat sheets joined by the bridge.

Einstein and Rosen found a simple trick that allowed them to avoid the matter stress-energy tensor $T_{\mu\nu}$ in Einstein’s field equations (9). They inserted the negative of the Maxwell stress-energy tensor $T_{\mu\nu}$ in vacuum into (9), and obtained Eqs. (24) of similar form to the modified vacuum field equations (18). But (24) led to the new massless-charged particle.

As Einstein and Rosen gradually refined their ER bridge theory, they drew close connections between mass and electric charge of elementary particles in terms of types of bridges. They now fully understood that the electron or proton needs two bridges between the two flat sheets because the electron and proton have both mass and electric charge. But the electron and proton, each one of them, should appear as one “hole in space” instead of one point-singularity. They cannot be represented by two “holes in space,” i.e. two bridges between the two flat sheets. Einstein found this blatant contradiction in his ER bridge theory most disturbing. At around the same time, Einstein wrote that the law of motion of material particles amounts to the discovery of solutions of the field equations that contain several bridges (Einstein, 1936, p. 380).

5 The Einstein–Podolsky–Rosen (EPR) Argument

5.1 The EPR Paper

I conjecture that the “Russian Jew” referred to in the letter to Besso quoted above is probably Boris Podolsky and the “many-body problem” is most likely an allusion to the title of Heisenberg’s 1926 paper (Heisenberg, 1926): “Mehrkörperproblem und Resonanz in der Quantenmechanik” (Many-Body Problem and Resonance in Quantum mechanics), which discusses many-body systems in quantum theory in terms of the Bose–Einstein statistics, a gas of non-interacting particles.

It also seems likely that Einstein and Podolsky endeavored to solve the “many-bridge problem” using Heisenberg’s many-body paper. I conjecture that this attempt was a last resort to save the ER bridge theory. But it failed to yield solutions that represent electrons and protons. The unfortunate conclusion was that this was yet another failure to describe the whole field without introducing singularities.

The EPR thought experiment has been discussed many times in the literature. It deals with a quantum mechanical system comprising two partial systems A and B that are in interaction with each other only during a limited time. Let there be given the wave function $\psi(x_1, x_2)$ before their interaction. After the interaction, the system A is separated from B. The Schrödinger equation supplies the ψ-function for the two partial systems A and B after the interaction has taken place. However, we cannot calculate the wave function of either particle A or particle B alone. This, according to quantum mechanics, can be done only with the help of measurements. Let us now determine the physical condition of the partial system A as completely as possible by measurements. Quantum mechanics then allows us to determine the ψ-function of the partial system B from the measurements made and from the ψ-function of the total system. This determination, however, gives a result that depends upon which one of the values specifying the physical condition of A has been measured (position or momentum).

In early March 1935, Einstein et al. (1935) (EPR) submitted the manuscript of the EPR paper, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" to *Physical Review* where it was published on May 15, 1935. They end their paper by saying that they are forced to conclude that the quantum mechanical description of physical reality given by wavefunctions is not complete (Einstein et al., 1935, p. 780).

I am not going to discuss the EPR paper itself or responses to it because these are out of the scope of this paper. I only quote Abraham Pais saying (Pais, 1979, p. 904): "It should be stressed that this paper contains neither a paradox nor any flaw of logic. It simply concludes that objective reality is incompatible with the assumption that quantum mechanics is complete."

A year later, in a paper of March 1936, entitled "Physics and Reality," Einstein discussed "the paradox recently demonstrated by myself and two collaborators [Podolsky and Rosen], and which relates to the following problem." Just as he had written in private in 1935 to Erwin Schrödinger and Karl Popper, Einstein argued that the most immediate paradox is that: "Since there can be only one physical state of B after the interaction which cannot reasonably be considered to depend on the particular measurement we perform on the system A separated from B it may be concluded that the ψ-function is not unambiguously coordinated to the physical state. This coordination of several ψ-functions to the same physical state of system B shows again that the ψ-function cannot be interpreted as a (complete) description of a physical state of a single system" (Einstein, 1936, p. 376).

Thus, during 1935–1936, the non-separability of A and B was Einstein's most fundamental problem with quantum mechanics (Howard, 1985).

5.2 Two-Body Problem and Contiguity

In 1946, Einstein invoked a stricter principle, which, in 1948, he called the principle of contiguity (*Prinzip der Nahewirkung*), and which he linked to the field theory (A. Einstein to M. Born, April 5, 1948, (Einstein and Born, 1969, Letter 88), Manuscript, 226, "Quantum Mechanics and Reality"; (Einstein, 1948, pp. 321–323):

> The following idea characterizes the relative independence of objects far apart in space (A and B): external influence on A has no direct influence on B; this is known as the 'principle of contiguity', which is used consistently only in the field theory. [...]
>
> I now make the assertion that the interpretation of quantum mechanics [...] is not consistent with principle II [the principle of contiguity]. Let us consider a physical system S_{12}, which consists of two part-systems S_1 and S_2 [... they interact and are described by ψ_{12} ...]. At time t let the two part-systems be separated from each other in space [...].

In his *Autobiographical Notes*, Einstein presented yet another version of the EPR argument and concluded:

> For the same situation of S_2 it is possible therefore to find, according to one's choice, different types of ψ-function. (One can escape from this conclusion only by either assuming that the measurement of S_1 (telepathically) changes the real situation of S_2 or by denying independent real situations as such to things which are spatially separated from each other. Both alternatives appear to me entirely unacceptable) (Einstein, 1949, p. 80).

One can solve the EPR paradox "only by either" telepathy "or by" violating contiguity and separability. Einstein could certainly not accept telepathy. That is due to the fact that he spoke about "the real situation of" S_2. Thus, the EPR argument was still riven with paradoxes. He could not violate contiguity and separability either, as he explained in the 1948 piece. And again, the EPR argument was riven with contradictions.

In a much-quoted letter of March 3, 1947 from Einstein to Max Born, the 1946 "telepathically" became "spooky actions at a distance": "the theory cannot be reconciled with the idea that physics should represent a reality in time and space, free from spooky actions at a distance" (A. Einstein to M. Born, March 3, 1947, Einstein and Born, 1969, letter 84). Einstein could not accept spooky action at a distance between two particles. It seems most likely that in 1946–1947 Einstein retained contiguity and separability because he treated the EPR pair of particles as if it were a two-body problem in the field theory.

6 Concluding Remarks

It seems likely that during winter-spring 1935 Einstein, Podolsky, and Rosen had been working on two problems: the ER bridges and the EPR argument. First, Einstein and Rosen worked on the ER bridge theory and then they encountered a dead-end situation. They understood that the electron and proton needed two bridges between the two flat sheets because the electron and proton have both mass and electric charge. Einstein found this blatant contradiction in his ER bridge theory most disturbing. Second, Einstein set to solve this "many-bridge problem" together with Podolsky, as he told his friend Besso. But with regard to the EPR argument, after 1935, Einstein presented his own formulations of that argument, which may rightly be called Einstein's formulations of the EPR paradox. Stubbornly insisting on separability and contiguity, even at the cost of giving up the requirement of simplicity and complicating the explanation of the EPR thought experiment to the point of creating a paradox, seems to support my conjecture that Einstein treated the EPR pair of particles as a two-body problem in the field theory.

Acknowledgments This work is supported by ERC advanced grant 834735. I would like to thank Dennis Lehmkuhl for providing me with enlightening comments in the conference, "Thinking about Space and Time: 100 years of Applying and Interpreting General Relativity", Bern (Switzerland), September 12–14, 2017. I would also like to thank Tilman Sauer for the enlightening comments and suggestions and thank the anonymous referee for the helpful comments on my manuscript.

References

Bach, Rudolf and Weyl, Hermann (1922). "Neue Lösungen der Einsteinschen Gravitationsgleichungen. B. Explizite Aufstellung statischer axialsymmetrischer Felder. Mit einem Zusatz über das statische Zweikörper-Problem. Von H. Weyl." *Mathematische Zeitschrift* 13,134–145. First part by Bach, second part by Weyl, Hermann, "Das statische ZweikörperProblem." (142–145).

Chazy, Jean (1924). "Sur le champ de gravitation de deux masses fixes dans la théorie de la relativité." *Bulletin de la Société Mathématique de France* 52, 17–38.

Curzon, Harry E. J. (1924). "Cylindrical Solutions of Einstein's Gravitation Equations" *The London Mathematical Society. Proceedings* 23, 477–480.

van Dongen, Jeroen (2010). *Einstein's Unification*. Cambridge: Cambridge University Press.

Earman, John and Eisenstaedt, Jean (1999). "Einstein and Singularities." *Studies in History and Philosophy of Modern Physics* 30, 185–235.

Einstein, Albert (1936). "Physics and Reality." *Journal of the Franklin Institute* 221, 349–382.

Einstein, Albert. (1948). "Quanten-Mechanik und Wirklichkeit." *Dialectica* 2, 320–324.

Einstein, Albert. (1949). "Autobiographisches/Autobiographical Notes." In [Schilpp, Paul Arthur (ed.) (1949). *Albert Einstein: Philosopher-Scientist*. La Salle, IL: Open Court.].

Einstein, Albert and Besso, Michele (1972). *Correspondence 1903–1955*. Translated by Pierre Speziali, Paris: Hermann.

Einstein, Albert and Born, Max (1969). *Albert Einstein Max Born Briefwechsel 1916–1955*. Munich: Nymphenburger Verlagshandlung GmbH. The Born-Einstein Letters. Translation by Irene Born with commentaries by Max Born. Glasgow: Macmillan, 1971.

Einstein, Albert and Grommer, Jakob (1927). "Allgemeine Relativitätstheorie und Bewegungsgesetz." *Preussischen Akademie der Wissenschaften* (Berlin), *Physikalisch-mathematische Klasse. Sitzungsberichte*, 2–13. Reprinted with commentary in *The Collected Papers of Albert Einstein*. Vol. 15 *The Berlin Years: Writings and Correspondence, June 1925–May 1927*, ed. Kormos Buchwald, Diana *et al.*, Princeton: Princeton University Press, 2018, Doc. 443.

Einstein, Albert, Infeld, Leopold and Hoffmann, Banesh (1938). "The Gravitational Equations and the Problem of Motion." *Annals of Mathematics* 39, 65–100.

Einstein, Albert, Podolsky, Boris and Rosen, Nathan (1935). "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review* 47, 777–780.

Einstein, Albert and Rosen, Nathan (1935). "The Particle Problem in the General Theory of Relativity". *Physical Review* 48, 73–77.

Einstein, Albert and Rosen, Nathan (1936). "Two-Body Problem in General Relativity Theory." *Physical Review* 49, 404–405.

Havas, Peter (1993). "The General-Relativistic Two-Body Problem and the Einstein-Silberstein Controversy." In Earman, John, Janssen, Michel and Norton, John (eds.) (1993). *The Attraction of Gravitation: New Studies in the History of General Relativity*. Einstein Studies 5. Boston: Birkhäuser, 88–125.

Heisenberg, Werner (1926). “Mehrkörperproblem und Resonanz in der Quantenmechanik.” *Zeitschrift für Physik* 38, 411–426.
Howard, Don, “Einstein on Locality and Separability.” *Studies in History and Philosophy of Science* 16, 1985, 171–201.
Infeld, Leopold (1941). *Quest: An Autobiography*. New-York: Doubleday, Doran & Company, Incorporated. Second edition, Rhode Island: American Mathematical Society, 1980.
Laurence, William I. “Einstein in Vast new Theory Links Atom and Stars in Unified System.” *The New York Times*, (Friday) July 5, 1935.
Levi-Civita, Tulio (1919). “Meccanica – ds^2 einsteiniani in campi newtoniani. VIII. Soluzioni binarie di Weyl.” *Rendiconti della Reale Accademia dei Lincei* 28, 3–13.
Maldacena, Juan and Susskind, Leonard (2013). “Cool Horizons for Entangled Black Holes.” *Fortschritte der Physik* 61, 781–811.
Pais, Abraham (1979). “Einstein and the quantum theory.” *Reviews of Modern Physics* 51, 863–914.
Rosen, Nathan (1949). “A Particle at Rest in a Static Gravitational Field.” *Reviews of Modern Physics* 21, 503–505.
Sauer, Tilman (2014). “Einstein’s Unified Field theory Program.” In Janssen, Michel and Lehner, Christoph (eds.) *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press, 2014, pp. 281–305.
Schilpp, Paul Arthur (ed.) (1949). *Albert Einstein: Philosopher-Scientist*. La Salle, IL: Open Court.
Silberstein, Ludwig (1936). “Two-Centers Solution of the Gravitational Field Equations, and the Need for a Reformed Theory of Matter” *Physical Review* 49, 268–270.
Weyl, Hermann (1917). “Zur Gravitationstheorie.” *Annalen der Physik*, 54, 117–145.
Weyl, Hermann. (1919). “Bemerkung über die axialsymmetrischen Lösungen der Einsteinschen Gravitationsgleichungen.” *Annalen der Physik* 59, 185–188.

A Raum with a View: Hermann Weyl and the Problem of Space

Neil Dewar and Joshua Eisenthal

Abstract A central issue in the philosophical debates over general relativity concerns the status of the metric field: should it be regarded as part of the background arena in which physical fields evolve, or as a physical field itself? In this paper, we approach this debate through its relationship to the so-called "Problem of Space": the problem of determining which abstract, mathematical geometries are candidate descriptions of physical space. In particular, we explore the way that Hermann Weyl tackled the Problem of Space in the wake of general relativity, and argue that Weyl's proposed solution reveals a "middle way" between bare-manifold and manifold-plus-metric accounts of spacetime.

1 Introduction

One of the central philosophical debates prompted by general relativity concerns the status of the metric field. A number of philosophers have argued that the metric field should no longer be regarded as part of the background arena in which physical fields evolve; it should be regarded as a physical field itself. Earman and Norton write, for example, that the metric tensor in general relativity 'incorporates the gravitational field and thus, like other physical fields, carries energy and momentum'.[1] Indeed, they baldly claim that according to general relativity 'geometric structures, such as the metric tensor, are clearly physical fields in spacetime'.[2] On

[1]Earman and Norton (1987, p. 519).

[2]Earman and Norton (1987, p. 519).

N. Dewar (✉)
Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, München, Germany
e-mail: neil.dewar@lrz.uni-muechen.de

J. Eisenthal
Department of Philosophy, University of Pittsburgh, Pittsburgh, PA, USA

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_6

such a view, spacetime itself—considered independently of matter—has no metrical properties, and the mathematical object that best represents spacetime is a bare topological manifold. As Rovelli puts the idea: 'the metric/gravitational field has acquired most, if not all, the attributes that have characterised matter (as opposed to spacetime) from Descartes to Feynman... it is perhaps more appropriate to reserve the expression *spacetime* for the differential manifold, and to use the expression *matter* for everything which is dynamical, carries energy and so on; namely all the fields *including the gravitational field*'.[3]

Others, however, have strongly resisted this view, arguing that the paradigm spatio-temporal properties are precisely the metrical properties. Thus Maudlin has written:

> ...qua differentiable manifold, abstracting from the metrical (and affine) structure, spacetime has none of the paradigm spatio-temporal properties. The light-cone structure is not defined; past and future cannot be distinguished; distance relations do not exist. Spatio-temporal structure is *metrical* structure.[4]

In resisting the adoption of a bare-manifold account of spacetime Maudlin has been joined by Hoefer, who has argued that the mere fact that the metric field appears to carry energy does not imply that it should be regarded as a physical field.[5] More specifically, Hoefer has argued that the complexities arising from the fact that gravitational stress-energy is represented by a *pseudo*-tensor cannot be merely brushed aside, and that there are good reasons to be sceptical of any quick inferences based on the existence of gravitational energy in this context.[6]

This debate—between *bare-manifold* and *manifold-plus-metric* accounts of spacetime—touches on the long-standing philosophical problem of how mathematics represents the world. In the context of geometry, the most immediate aspect of this problem is the 'Problem of Space': the problem of determining which abstract geometrical structures are candidate *physical* geometries, i.e. candidate descriptions of physical space. It is only since the advent of non-Euclidean geometries that this problem has emerged, or could even be stated. For most of its history, of course, geometry was just *Euclidean* geometry, understood as the systematic description of spatial structure ('the most ancient branch of physics', as Einstein once put it).[7] Hence it was only after the existence of non-Euclidean geometries was grudgingly accepted that it became possible to ask: *which* geometry actually describes space? By the end of the nineteenth century (as we summarise below) consensus formed around the following answer: the candidate physical geometries are the *constant curvature* geometries; the geometries in which congruence relations can represent the free mobility of rigid bodies.

[3] Rovelli (1997, pp. 193–194).

[4] Maudlin (1988, p. 87; emphasis added).

[5] Cf. (Hoefer, 1996).

[6] For further discussion of the controversies over gravitational energy, (see Curiel, 2019; Read, 2018; Dewar and Weatherall, 2018).

[7] Einstein (1921, p. 6).

However, this ‘classical’ solution to the Problem of Space was unequivocally undermined by general relativity—the theory of spacetime that employs precisely the kind of variably curved geometry that the philosophers of the nineteenth century thought they had ruled out. In this paper, we explore the new solution to the Problem of Space advanced by Hermann Weyl, drawing especially on the account Weyl gave of this problem in a series of lectures delivered in Barcelona in 1922.[8] Weyl had an exceptionally nuanced understanding of the novel conception of spacetime implicit in general relativity, and our concrete goal in what follows is to show that an important insight made available by Weyl's work is the unearthing of a ‘middle way’ between bare-manifold and manifold-plus-metric accounts of spacetime.

2 The Classical Solution

The classical solution to the Problem of Space is, in effect, the fruits of the cumulative effort of those who engaged with the problem in the second half of the nineteenth century. The formulation of the Problem of Space in this context is often referred to as the *Helmholtz* (or *Helmholtz/Lie*) *space problem*, but as Helmholtz himself pointed out it had already been treated substantially by Riemann. Helmholtz's results were rigorously reworked and extended by Sophus Lie, who brought to bear the full power of his theory of continuous transformation groups.[9] Poincaré also grappled with the Problem of Space, and although his philosophical stance differed significantly from Helmholtz's, he clearly regarded the problem as more-or-less solved once Lie had put Helmholtz's arguments on a sufficiently rigorous footing.[10]

The essential idea that emerged in this period was that the geometrical notion of congruence represented the possibility of the free mobility of rigid bodies. Treating such free mobility as a basic fact (and recognising the role that this fact seemed to play in the practice of measurement quite generally), a limited class of geometries could be specified as candidate descriptions of physical space. In particular, it was argued that only *constant curvature geometries* could represent the free mobility of rigid bodies because only these geometries have suitable congruence relations between geometric figures.

The close connection between congruence relations of geometrical figures and constant curvature was proved most rigorously by Lie, who considered the

[8]Weyl (1923a).

[9]For Riemann and Helmholtz, the problem at hand was one in which physics and mathematics necessarily intertwined. In contrast, Lie treated the problem as the purely mathematical one of rigorously characterising distinct classes of geometrical structures. The articulation of the range of mathematical possibilities was developed also by Klein, Clifford, and Killing: see Scholz (2016, p. 5), for a summary of this work.

[10]Poincaré remarked, ‘I differ from [Helmholtz and Lie] in one point only, but probably the difference is in the mode of expression only and at bottom we are completely in accord’ (Poincaré, 1898, p. 40).

properties of the real, finite-dimensional transformation groups that correspond to classical congruence relations, but it is also easy to understand the basic point intuitively. Helmholtz discussed the example of the non-constant curvature of an egg-shaped surface, observing that a figure such as a triangle would have different internal angles if it were drawn near the pointed end of the egg, compared to if it were drawn near the base.[11] Similarly, two circles with the same radius would not in general have the same circumference. Thus sliding any such figure up or down the surface would not be possible unless it were flexible enough to change its dimensions as it moved. In the general case, in any space of varying curvature (with zero degrees of symmetry), a truly rigid figure—one that could not alter its dimensions without breaking—would not be able to be moved at all.

In this way, Helmholtz, and Poincaré after him, argued that geometries of *non*-constant curvature are not feasible candidates for describing physical space *at all*. Their reasoning depended on the premise that constructing a physical geometry depends essentially on the use of rigid bodies. According to Helmholtz, this is a generalisation of the requirement that material measuring instruments (rulers, compasses, and the like) must maintain their dimensions if they are to fulfil their function.[12] Helmholtz argued that if there were no rigid bodies, and hence no way of comparing spatial magnitudes, it would not be possible to construct any kind of physical geometry. Hence variably curved geometries, which lack the relevant congruence structure, cannot provide a useful description of physical space (for even if we lived in such a space, so the thought goes, we would not be able to construct a geometrical description of it). Thus Helmholtz declared: 'all original spatial measurement depends on asserting congruence and therefore, the system of spatial measurement must presuppose the same conditions on which alone it is meaningful to assert congruence'.[13] Poincaré, for his part, said of variably curved geometries that they could 'never be anything but purely analytic, and they would not be susceptible to demonstrations analogous to those of Euclid'.[14]

This, then, was the classical solution to the Problem of Space: a geometrical structure could describe spatial structure only insofar as it could represent the free mobility of rigid bodies. Combined with Lie's work, this postulate of free mobility implied a clear demarcation of candidate physical geometries from two directions. First, candidate physical geometries could only be constant curvature geometries. Second, the metric function must satisfy a generalised Pythagorean Theorem, i.e. the element of length must be given by the square root of a quadratic differential form.

However, the development of relativity—particularly general relativity—broke the back of this purported solution to the Problem of Space. One obvious change due to relativity was the new way in which space and time were welded together into spacetime, but the more immediately significant change was actually the threat

[11] Helmholtz (1995, p. 231).
[12] Helmholtz (1995, p. 239).
[13] Helmholtz (2007, p. 49).
[14] Poincaré (1952, p. 103).

to the notion of a rigid body. Already in special relativity it became clear that perfectly rigid bodies were simply incompatible with the theory.[15] But it was only with the advent of general relativity that the classical solution to the Problem of Space was definitively undermined, for it is general relativity that employs precisely the variably curved geometrical structure that Helmholtz and Poincaré had exiled from the class of candidate physical geometries. As Scholz has put the matter, general relativity 'posed, of course, a much greater challenge to the characterisation of the Problem of Space. Free mobility of finitely extended rigid figures became meaningless in the general case.'[16]

3 Weyl's Problem of Space

In the wake of general relativity it is evident that the classical demarcation of candidate physical geometries is too restrictive—clearly, an adequate demarcation must include variably curved geometries too. This is the context in which Weyl sought to justify a new and broader conception of physical geometry. On Weyl's view, the possibility of describing physical space in geometrical terms depends only on the possibility of *infinitesimal* comparisons of lengths and angles. This still allows for the construction of practically rigid bodies, so that, as long as circumstances are not too hostile, we can still survey the meso-scale structure of space in the way that Helmholtz and others envisaged. But Weyl's solution also leaves room for the possibility of describing the geometry of a region of space encompassing such strongly varying gravitational fields that surveying it with rigid measuring instruments would be impossible.

An important upshot of Weyl's approach to the Problem of Space is a distinction between the *nature* of the metric field, on the one hand, and the *orientation* of the metric field from point to point, on the other. The former is what determines the relative lengths of vectors at an arbitrary point, whilst the latter is what determines the relative lengths and angles of finitely separated vectors. Weyl uses this distinction to attribute the local metric properties to space itself, whilst attributing the non-local metric properties to the contingent distribution of matter and energy. In brief, rather than starting with the classical postulate of free mobility, Weyl starts with what he calls the 'foundational fact of infinitesimal geometry'[17]—the idea that the notion of congruent transport (the transformation that preserves *length*) uniquely determines a notion of affine transport (the transformation that preserves *parallelism*). Weyl then proves that this foundational fact provides the basis for a new demarcation of physical geometries. According to Weyl, the candidate physical geometries are

[15] For a more detailed discussion of the impact of special relativity on the notion of rigid bodies, see Scholz (2016, pp. 6–8).

[16] Scholz (2016, p. 9).

[17] Weyl (1923b, p. 124), quoted at more length below.

just the geometries with a Pythagorean-Riemannian 'nature'; the geometries whose metrics have an *infinitesimal* Pythagorean form.

It seems that Weyl first became interested in the Problem of Space when he was called upon to edit Riemann's 1868 *Habilitationsvortrag* for republication, for which he also provided a commentary.[18] In the *Habilitationsvortrag*, Riemann gives the length of a line element as the square root of a quadratic form in the differentials, but remarks that this is merely the simplest case of a wider range of possibilities:

> The next simplest case would probably comprise those manifolds in which the line element may be expressed by the fourth root of a differential expression of the fourth degree. To be sure, the investigation of this more general kind would not require any essentially different principles, but it would be rather time-consuming and would shed relatively little new light on the theory of space (especially as the results are not geometrically expressible); I therefore restrict myself to those manifolds where the line element is expressed by the quadratic root of a differential expression of the second degree.[19]

Weyl expresses Riemann's observation as follows: if the interval at point P is expressed as a function of the differentials, $ds = f_P(dx_1, \dots, dx_n)$, then '$f_P$ will be required to be a homogeneous function of the first degree, in the sense that upon multiplication of the arguments dx_i by a common real proportionality factor ρ, the function f_P is multiplied by $|\rho|$.'[20] An example of such a function is the familiar Pythagorean function:

$$\sqrt{(dx_1)^2 + (dx_2)^2 + \cdots + (dx_n)^2} \tag{1}$$

Up to a choice of coordinates, any function given as a square root of some positive-definite quadratic form can be expressed in the form (1) (i.e. if f_P^2, at each point P, is a positive-definite quadratic form, then all the various f_P can be obtained from the function (1) by linear transformations of the variables). However, (1) is not the only homogeneous function of the first degree, and so the question arises: why use this function to define intervals, rather than any others? Riemann himself offered no satisfactory justification for why the expression for the square of the line element should be a quadratic form, and hence it was a signature achievement of the classical solution to the Problem of Space to show that, if every physical geometry must represent the free mobility of rigid bodies, every physical geometry must have a Pythagorean metric. But when general relativity undermined the classical solution to the Problem of Space, this justification for the Pythagorean form of the metric vanished with it.

[18] Riemann and Weyl (1919).

[19] Riemann and Weyl (1919, p. 9), Riemann and Jost (2013, p. 35).

[20] Riemann and Weyl (1919, p. 26). In modern terminology, a homogeneous function is one for which $f(\lambda v) = \lambda f(v)$, whereas a function with the property that Weyl describes (being such that $f(\lambda v) = |\lambda|v)$ is described as *absolutely* homogeneous. In what follows, we will use 'homogeneity' in the same manner as Weyl (i.e. as a term for what is now called absolute homogeneity).

Thus, if we follow Weyl and say that the *nature* of a (Weylian) metric is given by specifying the expression for the line element, then the problem at hand is to justify the Pythagorean nature in particular. Weyl's solution to this problem begins from his generalisation of Riemannian geometry.[21] In a Riemannian space, (M, g), where M is a manifold and g is a metric, consider two arbitrary tangent vectors, $v \in T_pM$ and $w \in T_qM$, for finitely separated points p and q. Whether or not these two vectors are parallel is not, in general, determined absolutely, but only relative to the choice of a path connecting p and q. This becomes most apparent if we introduce the concept of an *affine connection* following Levi-Civita.[22] An affine connection establishes the parallelism-facts amongst the vectors in infinitesimally separated tangent spaces; it is only in the special case of a *flat* affine connection, however, that these can be extended to establish absolute (i.e. path-independent) parallelism-facts between vectors in finitely separated tangent spaces.[23]

For Weyl, Levi-Civita's work presented a profound insight into the structure of Riemannian geometry. But he soon recognised that it pointed to a way in which Riemannian geometry was not as local as it could be. Although Riemannian geometry does away with an absolute notion of distant parallelism, it retains an absolute notion of distant *length-comparison*: there is always a definite answer to the question whether vectors $v \in T_pM$ and $w \in T_qM$ are the same length or not, even for finitely separated p and q. More generally, for any two tangent vectors, there is some fact of the matter about their lengths, and hence about their length-ratio.

In aiming for a truly local geometry, then, our first move should be to do away with the structure of Riemannian geometry which permits such comparisons. Consider a pair of *conformally equivalent* metrics on M: that is, metrics g_{ab} and g'_{ab} such that for some smooth, positive, nowhere-vanishing scalar field $\lambda : M \to \mathbb{R}$,

$$g'_{ab} = \lambda g_{ab} \, . \tag{2}$$

A *conformal structure* on M is an equivalence class of conformally equivalent metrics; a *conformal manifold* is a manifold equipped with a conformal structure.[24] As Weyl remarks, in a conformal geometry the inner product of two vectors (in the

[21] Weyl (1918a,b).

[22] Levi-Civita (1917).

[23] Both Levi-Civita and Weyl took the term 'affine connection' to refer exclusively to *symmetric* (i.e. torsion-free) affine connections. We will have cause, however, to consider nonsymmetric affine connections in Sect. 4, so we will use the term 'affine connection' to refer to the broader class of such connections (whether symmetric or nonsymmetric).

[24] There is a natural sense, using the language of category theory, in which a conformal manifold is less structured than a Riemannian manifold. Let the category of Riemannian manifolds have isometries as arrows (and Riemannian manifolds as objects), and let the category of conformal manifolds have conformal maps as arrows (and either conformal manifolds or Riemannian manifolds as objects). Then there is a functor from the category of Riemannian manifolds to the category of conformal manifolds which is faithful and essentially surjective, but not full: i.e. which 'forgets only structure', in the terminology of Baez and Shulman (2010, §2.4).

same tangent space) is ‘not absolute, but rather determined only up to an arbitrary non-zero proportionality factor’.[25]

Conformal manifolds do not permit distant length-comparisons: we can only determine the length-ratio between two vectors if they are drawn from the same tangent space. But in the passage to conformal manifolds, we have removed more structure than we wished to. For the analogy to parallelism (in Riemannian geometry) to hold, we do not want determinations of length-ratio between distant vectors to be *impossible*: we just want them to be relative to a choice of path. Equivalently, we want there to be (absolute) facts about the length-ratios of infinitesimally separated vectors, just not about the length-ratios of finitely separated vectors.

We must therefore restore the kind of structure that will let us make such comparisons, i.e. something analogous to an affine connection but for lengths rather than directions: ‘a *concept of transfer of the length-unit from a point P to its immediate neighbours*’.[26] Let us refer to such a standard of length-transfer as a *length connection*.[27] By transferring an (arbitrarily chosen) length-unit from one point to another, we can compare the lengths of vectors in the tangent spaces at the two points, and the length connection operates in such a way that this length-ratio is independent of the choice of unit. Thus, in Weylian geometry, length-comparisons of distant vectors are once again possible but—in general—only relative to a path. As is the case for affine connections, things may work out such that the value of the integral is path-independent; in this case, we say that the length connection is *flat*, and the geometry is that of a Riemannian manifold up to an arbitrary global choice of length-scale. We will refer to a conformal structure together with a length connection as a *Weylian metric*, and a manifold equipped with a Weylian metric as a *Weylian manifold*.

Returning to our main theme, the groundwork for Weyl’s solution to the Problem of Space was laid already in his 1919 discussion:

> Given the fundamental significance for the construction of geometry which, following recent investigations ..., attaches to the basic affine concept [*affine Grundbegriff*] of the infinitesimal parallel transport of a vector, the question in particular arises, whether the manifolds of the Pythagorean class of spaces [i.e. those in which the line element can be given in the Pythagorean form (1)] are the only ones which permit the establishment of this concept, and which correspondingly possess not only a metric, but also an affine connection. The answer is most likely affirmative, but a proof has so far not been rendered.[28]

As already noted, it is this idea which drives Weyl’s solution to the Problem of Space. More specifically, the key insight is that only the Pythagorean kind of spaces have the feature that they are associated with a *unique* concept of parallel transport.

[25] Weyl (1918b, p. 396).

[26] Weyl (1918b, p. 397).

[27] Note that this term, now standard in the literature, is not used by Weyl himself, who speaks instead of a *metrische Zusammenhang*, literally: ‘metrical connection’.

[28] Riemann and Weyl (1919, p. 27).

As is well-known, for a given Riemannian metric there is a unique symmetric affine connection compatible with it; compatible, that is, in the sense that parallel-transported vectors retain their length.[29] This result extends to Weylian manifolds: given a Weylian manifold, there is a unique compatible affine connection.[30]

This fact—the uniqueness of the affine connection, given the metrical structure—was greatly striking to Weyl, and something he put great emphasis on:

> And now we come to that fact, already anointed above as the *foundational fact of infinitesimal geometry*, which brings the construction of geometry to a wonderfully harmonious conclusion. In a metrical space, there is one and only one way to formulate the concept of parallel transport so that . . . this postulate is fulfilled: *upon parallel transport of a vector, the interval determined by it should also remain unchanged*. Thus, the principle of infinitesimal *interval- or length-transfer*, which underlies metrical geometry, automatically brings with it a principle of *direction transfer*; *a metrical space naturally carries an affine connection*.[31]

At the heart of Weyl's solution to the Problem of Space is a demonstration that, among the much broader class of spaces obtained by allowing the line element to be an arbitrary homogeneous function of the differentials, only the Pythagorean spaces (i.e. the Weylian manifolds) will satisfy the following condition: 'whatever quantitative configuration (within the scope of the nature of the metric) the metric field may have assumed, it invariably and uniquely determines the affine connection.'[32]

4 Weyl's Solution

We turn now to a reconstruction of (part of) Weyl's argument. We follow the treatment given in Weyl (1923a): the text of a series of lectures on the Problem of Space which Weyl gave in the spring of 1922 in Barcelona and Madrid. There are some significant differences between the way in which Weyl carries out the argument here, compared to the way it is presented in his other work;[33] moreover, this text has not been translated, and so we hope that the discussion here can help bring these ideas to a wider audience.

Weyl reaches his solution via a group-theoretic analysis. Given an n-dimensional Weylian manifold, of whatever nature, let us say that a linear automorphism g of the tangent space at P is *congruent* if it preserves the interval: that is, if for any vector ξ at P, $f_P(g(\xi)) = f_P(\xi)$. Since the composition of two congruent automorphisms will similarly be a congruent automorphism, as will the inverse of any congruent

[29] Malament (2012, Lemma 1.9.1).

[30] Folland (1970, Theorem 2).

[31] Weyl (1923b, p. 124).

[32] Weyl (1923a, pp. 46–47).

[33] In particular, it does not involve the so-called 'Postulate of Freedom': see Appendix 1 for discussion.

automorphism, the collection of all congruent automorphisms of P's tangent space will form a *group*: let us refer to this group as the *congruence group* at P. For example, if f_P is the Pythagorean function (1), then the congruence group will be the orthogonal group $O(n)$.[34]

Note that in general, the congruence group is not sufficient to determine the nature of the metric. For example, the four-norm and the six-norm on $\mathbb{R}^2$, i.e. the functions

$$||(x, y)||_4 := (x^4 + y^4)^{1/4} \tag{3}$$

and

$$||(x, y)||_6 := (x^6 + y^6)^{1/6} \tag{4}$$

respectively, have the same congruence group: the group consisting of right-angle rotations and reflections about the x- and y-axes.[35] Nevertheless, if the tangent space at P has $O(n)$ as its congruence group, then the nature in question is Pythagorean (i.e. it is given by a positive-definite quadratic form) at P. So if Weyl can find conditions which guarantee that each tangent space has $O(n)$ as its congruence group, then he will have succeeded in showing that only manifolds with a Pythagorean nature satisfy those conditions.[36]

In the seventh of his Barcelona lectures, Weyl argues that if we postulate uniqueness of parallel transport, then the congruence group's Lie algebra (i.e. the collection of infinitesimal congruent automorphisms)[37] must be of dimension $n(n-1)/2$, and must satisfy a certain kind of antisymmetry condition (stated in more detail below). In the eighth and final lecture, Weyl sketches a proof that these conditions entail that the congruence group is the orthogonal group $O(n)$, and shows this by explicit calculation for the case $n = 2$; a complete proof (for the case of arbitrary dimensions) is provided in the appendices.[38]

[34] Weyl uses the term 'Drehungsgruppe' for what we are calling the congruence group, which would more literally be translated as 'rotation group'; however (as Coleman and Korté, 2001 note), the term 'rotation group' is nowadays almost exclusively used to refer to the groups $O(n)$ or $SO(n)$.

[35] Coleman and Korté (2001) make the same observation.

[36] Coleman and Korté (2001) castigate much of the literature for failing to appreciate that Weyl's task concerned singling out the Pythagorean nature from the broader class of possible natures for the metric, not that of singling out $O(n)$ from the broader class of congruence groups (see, in particular, Coleman and Korté, 2001, §§4.6–4.7). However, it is not clear to us that this difference is as significant as they suggest, given that singling out the desired congruence group is a sufficient condition for singling out the desired nature of the metric.

[37] Weyl does not use the term 'Lie algebra', but he notes that a collection of infinitesimal linear operations will form a linear family closed under the Lie bracket (again, not named as such): (see Weyl, 1923a, p. 50).

[38] A reconstruction of this second part of Weyl's solution, though certainly of value, is beyond the scope of this paper. The aspect of Weyl's proof is notoriously involved: Weyl describes himself

Weyl begins his argument by discussing the relationship between the metrical and the affine structure, i.e. between the transport of lengths and the transport of vectors. Recall that the length connection provides us with a unique way of transferring a length-unit from a point P to another point P_* in P's immediate neighbourhood, and hence of determining whether vectors ξ at P and ξ_* at P_* are the same length. In light of this, let us follow Weyl by saying that a linear isomorphism $\xi \in T_P M \to \xi_* \in T_{P_*} M$ is a *congruent transport* just in case it preserves the lengths of vectors: i.e. if the length of ξ relative to a given length-unit at P is the same as the length of ξ_*, relative to the transferral of that length-unit (using the length connection). Note that this is independent of the length-unit chosen at P.

At this stage we are restricting our attention to *infinitesimal* congruent transports: choosing some coordinate system around P, we are interested in those congruent transports $\xi \mapsto \xi_*$ for which

$$\xi_*^i = \xi^i + d\xi^i \tag{5}$$

Since congruent transports are linear, it follows that $d\xi^i = -\Lambda^i_k \xi^k$.[39] Let us say that the Λ^i_k are the *coefficients* of the congruent transport. Now, without yet specifying the dimensionality of the manifold, suppose that P_* is at the point $(\varepsilon, 0, 0, \ldots, 0)$. If we let Λ^i_{k1} be the coefficients of the congruent transport from P to P_*, and Λ^i_{k2} be the coefficients of a congruent transport from P to the point $(0, \varepsilon, 0, \ldots, 0)$, then (for any $\alpha, \beta \in \mathbb{R}$) we can show that $\alpha\Lambda^i_{k1} + \beta\Lambda^i_{k2}$ are the coefficients of a congruent transport from P to the point at $(\alpha\varepsilon, \beta\varepsilon, 0, \ldots, 0)$. Hence, once n congruent transports Λ^i_{kr} have been chosen, there is uniquely fixed a congruent transport to any point in P's infinitesimal neighbourhood. As Weyl puts it:

> ...the formula,
>
> $$d\xi^i = -\Lambda^i_{kr}\xi^k (dx)^r \tag{6}$$
>
> supplies a *system of infinitesimal congruent transports* to the totality of points $P' = (dx^1, dx^2, \ldots, dx^n)$ of the neighbourhood of P.[40]

Weyl's argument then runs as follows. Take a point P_0 in our manifold M, and introduce some coordinate system around it. Take as given the congruence group G_0 at P_0 (but not the congruence group at any other point).[41] Now let Λ^i_{kr} be an *arbitrary* collection of n^3 numbers. For every point P in the infinitesimal

as first having worked it out 'not through contemplation of the sense of the above-mentioned conditions, but rather only through mathematical acrobatics' (Weyl, 1922b, p. 120).

[39] The minus sign is a matter of convention.

[40] Weyl (1923a, p. 48). Note that we have slightly altered Weyl's notation to fit with that of this essay.

[41] That is, we do not fix the *action* of the congruence group at other points; we know, from the fact that the nature of the metric is everywhere the same, that the congruence group at any other point will be isomorphic to G_0.

neighbourhood of P_0, we can define a linear isomorphism from $T_{P_0}M$ to $T_P M$ by (6). Since every vector determines a length-unit (the one according to which that vector is of unit length), we can use this to define a length-unit transfer from P_0 to P, i.e. a length connection. Moreover, if we let the congruence group G at P be defined as the image of G_0 under this linear isomorphism, then the isomorphism will be congruence-preserving, and hence Λ will represent a system of infinitesimal congruent transports.

Now that we have a length connection and congruence groups on the infinitesimal neighbourhood of P_0, *any* infinitesimal congruent transport from P_0 to P may be obtained from a given such transport by composition with some infinitesimal congruent transformation.[42] Consequently, any *system* of infinitesimal congruent transports may be obtained from our original system (that encoded by the Λ^i_{jk}) by specifying n such infinitesimal congruent transformations: one for each of the n linearly independent coordinate displacements dx^r. If we let A^i_{kr} be the infinitesimal congruent transformation associated with dx^r, then the action of such a transformation on an arbitrary vector ξ at P is given (in terms of components) by:

$$\xi^i \mapsto \xi^i - A^i_{kr}\xi^k(dx)^r \tag{7}$$

It is at this point that we impose the postulate mentioned above, 'that among all these systems of infinitesimal congruent transports, a unique one is to be found which is simultaneously a possible system of parallel displacement'.[43] It follows that there is a unique array of A^i_{kr} which will bring about such a system of parallel transport. Using the Christoffel symbols now ubiquitous in general relativity, a parallel transport can be represented by Γ^i_{jk}, subject only to the symmetry requirement that $\Gamma^i_{jk} = \Gamma^i_{kj}$. We can then state Weyl's postulate as follows: given any Λ^i_{jk}, there is a unique system of parallel transport Γ^i_{jk} and a unique system of infinitesimal congruent transformations A^i_{jk} such that

$$\Lambda^i_{jk} = \Gamma^i_{jk} - A^i_{jk} \tag{8}$$

From this, Weyl proceeds to draw the following conclusions. First, if the dimensionality of the congruence group G (and hence, of its Lie algebra $\mathfrak{g}$) is N, then since every Λ^i_{jk} (with n^3 independent parameters) corresponds to a unique Γ^i_{jk} ($n^2(n+1)/2$ independent parameters) and A^i_{jk} (nN independent parameters), then $n^3 = \frac{n^2(n+1)}{2} + nN$; that is,

$$N = \frac{n(n-1)}{2} \tag{9}$$

[42] That is, an element of G's Lie algebra.

[43] Weyl (1923a, p. 49).

Second, note that if $A^i_{jk} = A^i_{kj}$, then Λ^i_{jk} will represent a system of parallel transport, which must therefore be identical with that represented by Γ^i_{jk}—from which it follows that $A^i_{jk} = 0$. So the family of the A^i_{jk} has the feature that they are symmetric ($A^i_{jk} = A^i_{kj}$) only if they all vanish. These two conclusions are the conditions on the congruence group's Lie algebra which—as discussed above—will lead us to the conclusion that the congruence group must be the orthogonal group (by an argument that we forbear from reconstructing here).

It bears emphasising how striking Weyl's achievement here is. Not only has he shown how the Problem of Space may be reframed in the light of general relativity; he has also shown that a satisfactory solution may be arrived at by the requirement that one's standard of length-comparison uniquely fixes one's standard of direction-comparison. It should also be stressed that Weyl's analysis is independent of his unorthodox geometrical background, insofar as Riemannian geometry is (up to an arbitrary choice of global scale) a special case of Weylian geometry.

In the wake of general relativity, it is evident that something fundamental has shifted in the implicit assumptions built into our practices of describing space geometrically. Weyl's solution to the Problem of Space offers an insight into precisely this fundamental shift. For Weyl, the possibility of describing space in geometrical terms depends only on the possibility of an idealised observer at a point, 'freely mobile' in the sense of being free to rotate at that point and start moving in any direction. Although such an observer can compare the dimensions of (infinitesimal) bodies in her immediate vicinity—that is determined only by the *nature* of space itself—what she might go on to discover about the larger-scale structure of space as she explores larger regions of it is left maximally unconstrained.[44] Weyl sums up this new conception of space, implicit in general relativity as he understood it, with the following vivid metaphor:

> Euclidean space may be compared to a crystal, built up of uniform unchangeable atoms in the regular and rigid unchangeable arrangement of a lattice; Riemannian space to a liquid, consisting of the same indiscernible unchangeable atoms, whose arrangement and orientation, however, are mobile and yielding to forces acting upon them.[45]

5 The Status of the Metric Field

Let us return to a consideration of what Weyl's work can contribute to the relatively recent debate over the status of the metric field. Recall that, on the one hand, because in general relativity the metrical field incorporates the gravitational field, some (including Earman, Norton, and Rovelli) have argued that the metric tensor

[44] For the curious reader, it is this Weylian notion of an idealised observer, free to rotate and move in any direction, that inspired the title for this paper. (Thanks to Stephen Mackereth for the reminder that this title was, in fact, his idea.)

[45] Weyl (2009, p. 88).

should be regarded as representing a physical field, akin to the electromagnetic field. On such a view, it is the bare topological manifold, absent any metrical properties, that should be regarded as representing spacetime itself. On the other hand, others (including Maudlin and Hoefer) have argued that metrical structure remains at the heart of paradigmatic spatio-temporal structure.

In considering how Weyl himself might have responded to this debate, the following statement seems unequivocal:

> ...it is not correct to say that space or the world is in itself, prior to any material content, merely a formless continuous manifold in the sense of analysis situs; the *nature* of the metric is peculiar to it in itself, only the mutual orientation of the metrics at various points is contingent, a posteriori and dependent on the material content.[46]

Thus, Weyl contrasts the *nature* of the metric—which, as we have seen, has its character fixed by the relationship between affine and metrical structure—with the *orientation* of the metric, encapsulating the remaining degrees of metrical freedom:

> Thus one sees how the nature of the metric can be the same at every location, even while its quantitative determination, the—so to speak—mutual orientation of the metric at different points, is still very changeable and capable of continuously varying configurations. Thus, from this standpoint, the *a prioristic essence* of space (defined by the nature of the metric) ...is divorced from the mutual *orientation* of the metric at the different points, which is a posteriori, i.e., contingent and naturally dependent on material content ...[47]

For present purposes, Weyl's distinction between the a priori and a posteriori serves to indicate the metrical properties that he attributes to space itself as contrasted with the metrical properties he attributes to the particular distribution of matter and energy. The fact that Weyl explicitly states that empty space is *not* 'merely a formless continuous manifold' would seem to place him squarely against the view advanced by Earman, Norton, and Rovelli. But in fact Weyl's analysis allows for a distinction that none of the more contemporary protagonists have in view whilst capturing motivations from both sides. On the one hand, there is the awkwardness of regarding the dynamical aspects of the metric as attributable to space itself; on the other hand, there is the fact that a bare manifold seems genuinely insufficient to represent anything we would recognise as space. But Weyl's distinction between the *nature* and *orientation* of the metric field provides a way to retain the idea that space is intrinsically metrical without thereby being forced to attribute all the dynamical aspects of the metric field to space itself.[48] Weyl thus provides a "middle way" between bare-manifold and manifold-plus-metric accounts of spacetime, arguing that only the *local* metrical properties—properties which are independent of the variable distribution of matter and energy—are attributable to space itself.

---

[46] Weyl (1922b, p. 117).

[47] Weyl (1922a, p. 216).

[48] One question we are left with here is whether the 'mutual orientation of the metric from point to point' should be regarded as in some sense akin to a physical field, and, if so, how. See Appendix 2 for a (partial) attempt at an answer.

One thing that emerges from the debate over the status of the metric field is that, in the wake of general relativity, we lack a principled means of identifying which mathematical structures represent features of space. By contrast, this was something that the figures of the nineteenth century had available to them: the classical solution to the Problem of Space provided a justification for why a particular collection of mathematical structures (the constant curvature geometries) could play this particular representational role. Once the classical solution became untenable, however, this justification went with it. Weyl's new solution to the Problem of Space thus offers a new justification for why an enlarged collection of mathematical structures—differential manifolds equipped with an infinitesimal Pythagorean-Riemannian metric—are candidate descriptions of physical space.

Throughout this paper, we have been treating Weyl's solution to the Problem of Space independently of his broader philosophical commitments. The fact that it is possible to do so points to the fundamental nature of the Problem itself. This is evident from the fact that, philosophical differences notwithstanding, there was broad *agreement* on the classical solution to the Problem of Space prior to general relativity. Poincaré, for example, could accept Helmholtz's solution to the Problem of Space whilst disagreeing with Helmholtz's claim that the value of the curvature of space would be determined by experiment. (For Poincaré, famously, the choice amongst constant curvature geometries was a matter of pure convention.) In a similar way, it is open to us to accept Weyl's new solution to the Problem of Space (and the insight into the conception of space implicit in general relativity that Weyl offers) independently of Weyl's own broader philosophical commitments.[49]

Weyl's argument seeks to show that reflection upon the concept of (physical) metrical structure—in particular, upon the required relationship between metrical structure and affine structure—provides a justification for the Pythagorean nature of the metric. This provides a different argument for regarding the metric as encoding spatial structure, beyond merely noting that certain 'paradigmatically spatial' properties depend upon it. In Weyl's analysis the *sine qua non* of physical geometry is that it realises a concordance between parallelism and congruence, and so the physical geometries are those whose infinitesimal metrical structures uniquely determine affine structures over finite distances.[50] It is with this kind of insight in view that we urge that engaging with the Problem of Space remains important, not merely as providing a different answer to the question of *what* represents space, but rather as a means of shedding light on the question of *what it is* to represent space.

[49]For a detailed discussion of the philosophical commitments framing Weyl's approach to the Problem of Space, especially Weyl's interest in Husserlian phenomenology, see Ryckman (2005, §6).

[50]It is worth noting that this requirement also ensures that *inertial structure* can be identified unambiguously: this suggests a connection between Weyl's work and Knox's analysis of spacetime as whatever plays the functional role of determining inertial frames (Knox, 2017).

Acknowledgments Our thanks to audiences at the 2015 Irvine-Pittsburgh-Princeton conference in Princeton, the 2015 meeting of the British Society for the Philosophy of Science in Manchester, and the 2019 Early Career Workshop in History and Philosophy of Physics in Pittsburgh. Thanks also to Gal Ben Porath, Stephen Mackereth, and an anonymous referee for their help in improving this paper. Special thanks to John Dougherty for contributing to this project across its various iterations and in particular for his detailed comments on an earlier draft.

Appendix 1: The 'Postulate of Freedom'

In our reconstruction of Weyl's solution, we founded Weyl's argument on only one postulate, rather than two. In this, we diverge from the more standard view in the literature, namely that Weyl's proof is based upon both a Postulate of Coherence (essentially the uniqueness postulate used in our discussion), *and* a Postulate of Freedom:[51] the postulate that any arbitrary collection Λ^i_{kr} of n^3 numbers represents a possible system of congruent transport. Now, this claim features in our reconstructions of Weyl's proof—but as an observation about the geometry, not as an independently stated postulate. It is widely accepted that this is the correct *logical* status to give to the Postulate of Freedom: since the work of Scheibe (1957), it is standard to describe the (so-called) Postulate of Freedom as a theorem rather than a postulate.[52] But what of its historical status? That is, did Weyl consider this to be a theorem or an independent postulate?

We submit that on this matter, Weyl changed his mind—but that the version given here represents his more mature view. Our reconstruction above follows the line of thought given in Weyl (1923a), in which Weyl only ever refers to *one* postulate (not two), in which the 'freedom of the metric' features as a preliminary observation, not a positive statement:

> I come now to the *synthetic part* in the Kantian sense. It is necessary to precisely formulate the postulate suggested earlier, which shall determine the kind of rotation group which is characteristic for the actual world. First, the freedom guaranteed to it! The free deformability of the metrical field is available to such a degree that for a given rotation group at P_0, the metrical connection of this point with the points in its surroundings can always be so formed that the equation
>
> $$d\xi^i = -\Lambda^i_{kr}\xi^k(dx)^r$$
>
> *with arbitrarily given coefficients* Λ^i_{kr}, represents a system of infinitesimal congruent transport of the tangent space at P_0. Second, the positive part of the postulate: however this metrical connection of P_0 with the points in its surroundings may have been formed, among the possible systems of parallel displacements of the tangent space there is always

[51] See, for instance, Hawkins (2000), Scholz (2001), Coleman and Korté (2001).

[52] See, e.g. Scheibe (1988), Scholz (2001, p. 92). Coleman and Korté (2001) are an exception, seemingly because of how strongly they reject Scheibe's interpretation of the Problem of Space.

> a unique one which is simultaneously a system of infinitesimal congruent transport; the metrical connection uniquely determines the affine.[53]

This is in sharp contrast to the version of this argument given elsewhere by Weyl, where—indeed—the postulates of freedom and coherence are distinguished from one another and separately stated. But that treatment represents an earlier phase of Weyl's thought: this discussion occurs first in the fourth edition of *Raum, Zeit, Materie* (Weyl, 1921), which he had completed by November 1920 (the date given in the preface, wherein he refers to 'a deeper group-theoretic formulation' of the Problem of Space as one of the major changes from the third edition); here, he succeeded in deriving the conditions on the Lie algebra, but not in showing that these conditions sufficed to uniquely identify the orthogonal group (except in the cases of two or three dimensions). A few months later, he had succeeded in completing the proof for the case of arbitrary dimension: the proof was submitted in April 1921 (published as Weyl, 1922b), and Weyl discussed the result in a lecture of September 1921 (the text of which was published as Weyl, 1922a).[54] All of these papers treat the Postulate of Freedom as a separate postulate, but all of them are prior to Weyl's delivery of the Spanish lectures (in spring 1922).

That said, it is true that Weyl also distinguished the two postulates in the *fifth* edition of *Raum, Zeit, Materie*, which was completed in autumn 1922. However, he finished preparing the *text* of the lectures (involving their translation from French and Castilian back into German, and the inclusion of appendices completing the proof) in April 1923.[55] Moreover, he appears to consider this text to supersede both (the treatment of the Problem of space in) the fifth edition, and the paper (Weyl, 1922b):

> I think of this little monograph primarily as an expansion of the book "Raum, Zeit, Materie" (5th ed., Julius Springer 1923). The deeper understanding of the problem of space, drawing on group theory, was there only briefly touched upon . . . ; that is made good upon here. . . .
>
> For the inclusion of a complete proof of that main group-theoretic result, which the problem of space leads us onto, I decided in the first instance . . . to simplify the first proof (Mathematische Zeitschrift **12**, p. 114) to a great extent.

For these reasons, it seems to us appropriate to take the one-postulate version of the argument as more reflective of Weyl's considered opinion.

Incidentally, there is an interesting question of the extent to which the published 1923 text reflects the lectures as they were delivered: in his preface, Weyl refers to the text as 'containing [the lectures] almost verbatim', but also remarks that the eighth lecture 'has had to undergo a sweeping revision'. Light could perhaps be shed on this matter if the projected Catalan version of the lectures had been published by the Institut d'Estudis Catalans, as was apparently planned (and happened for similar invited lecturers, e.g. Levi-Civita, 1922); however, it appears that this never came

---

[53] Weyl (1923a, p. 49).

[54] This timeline is based upon (Hawkins, 2000, §11.2).

[55] Weyl (1923a, p. III).

to fruition.[56] The document list for the Weyl archive at the ETH Zurich[57] does not appear to list notes for these lectures; it does, however, mention correspondence between Weyl and the Institut d'Estudis Catalans, and between Weyl and Esteban Terradas (who had invited Weyl to Barcelona, and who edited the series in which the book would have appeared), which could perhaps offer an account of why the Catalan book was never published.

Appendix 2: The Nature of the Metric as Background Object

As discussed in Sect. 3, the kind of homogeneous function used to express the norm at each tangent space specifies the nature of the metric: that is, if the interval is expressed in some coordinate system by the homogeneous function f, then the metric has nature (f) (where (f) is the equivalence class of f under arbitrary linear transformations). How do we capture the nature of a metric in more coordinate-free terms?

To this end, let us start by considering what it would be for two conformal manifolds to be of the *same* nature. If two conformal manifolds both have nature (f), then for any points $p \in M$ and $p' \in M'$, there exist coordinate systems $U \to \mathbb{R}^n$ and $U' \to \mathbb{R}^n$ (where $p \in U$ and $p' \in U'$) such that the intervals at p and p' are both expressed by f. But it then follows that there is a linear bijection $h : T_pM \to T_{p'}M'$ which preserves congruence. Conversely, if there exists such a map h, then given a coordinate system around p in which f expresses the interval at p, we can push forward the coordinate basis on T_pM to $T_{p'}M'$ under h, and then find a coordinate system around p' for which the pushed-forward basis is the coordinate basis; by construction, this basis will be one relative to which the interval at p' is expressed by f. Thus, two metrics are of the same nature if and only if there is a congruence-preserving map from any tangent space in one to any tangent space of the other.

Following this line of thought, we can represent the nature of the metric as follows. Suppose that we begin with an n-dimensional manifold M. Consider a fibre bundle which is isomorphic, *qua* vector bundle, to the tangent bundle TM; this bundle is distinguished from the tangent bundle by the fact that there is no privileged identification of points in the bundle with directional derivatives on the manifold. Now let W be the result of equipping this fibre bundle with a faithful action of $O(n)$ on each fibre (so $O(n)$ is the structure group of W). Let us (tendentiously) say that W is a *natured manifold*. Any conformal manifold can be regarded as a natured manifold, by letting the action of $O(n)$ on any tangent space be the group of congruence-preserving maps; and by the argument above, two conformal manifolds

[56]There is no reference to such a volume in the Weyl bibliography of Newman (1957), nor in the catalog of the Institut d'Estudis Catalan (Institut d'Estudis Catalans, 1997).

[57]Handschriften und Autographen der ETH-Bibliothek (1995).

will be isomorphic *qua* natured manifolds just in case they are diffeomorphic and of the same nature. So the name is appropriate, and we can think of W as a manifold equipped with a nature.

To turn W into a conformal manifold, we need to specify (in a smoothly varying fashion) a pointwise linear bijection $e : W \to TM$. Such a field e is known as a *solder form* or *vielbein*: in the context of general relativity, it is often called a tetrad field.[58] As before, to move from a conformal manifold to a Weyl manifold, we require also a length connection. Thus, on this reading, the nature of the metric is expressed in the congruence-structure of an auxiliary internal bundle; the orientation is captured by a tetrad field and a length connection.

We are not the only ones to have suggested that unpacking Weyl's distinction between nature and orientation might be associated with the vielbein formalism: for example, Ryckman makes the following suggestive remarks:

> The coordinate systems of the local group are defined in the tangent space over each point: each one is an orthonormal frame called a *tetrad* which is free to rotate independently of the tetrads over the other points. ... The effects of gravity are "restored" by a connection (here, the gravitational potential) which reconciles the local laws on various points. This is the distinction pointed to by Weyl in a purely mathematical context between the *nature* and the *orientation* of the metric ...[59]

This suggests, in contrast to our reading, that the nature of the metric be identified with the *tetrad field* (together with the internal metrical structure), rather than just the internal metrical structure: this would amount to identifying the nature of the metric with the conformal structure, and the orientation with the (length) connection.

In defence of our reading, we note not only that it is the most natural fit with Weyl's definition of nature in terms of equivalence classes of homogeneous functions, but also that it fits best with Weyl's insistence on the fixed and a priori character of the nature. We touched upon this already, but it is helpful to consider how Weyl introduces the distinction between nature and orientation in the 5th edition of *Raum-Zeit-Materie*:

> The *nature* of the metric signifies the aprioristic essence of space in its metrical aspect; it is *one*, thus it is also absolutely determined and does not partake of the irrevocable vagueness of those which occupy a variable place in a continuous scale. What is not determined through the essence of space, but rather is a posteriori (i.e. contingent, intrinsically free, and capable of arbitrary virtual changes), is the mutual *orientation* of the metrics at different points ...[60]

[58] See Weatherall (2016) for discussion of solder forms in general relativity.

[59] Ryckman (1999, p. 596). It may be noted that Ryckman identifies a tetrad as a frame field, i.e., a choice of basis at each point, rather than as a pointwise linear isomorphism between an internal bundle and the tangent bundle. The two definitions are equivalent if a preferred frame field for the internal bundle is chosen.

[60] Weyl (1923b, pp. 102–103).

With regard to this distinction, it is surely more natural to place conformal structure on the a posteriori side, given that Weylian manifolds (including Riemannian manifolds) can have very different conformal structures—as Weyl was surely aware, having introduced the conformally invariant Weyl curvature in Weyl (1918b). By contrast, the nature as we have characterised it is a local invariant in the following sense: given any two Weylian manifolds, expressed in vielbein form as $(\pi : W \to M, e)$ and $(\pi' : W' \to M', e')$, for any $p \in M$ and $p' \in M'$, there are neighbourhoods $U \ni p$ and $U' \ni p'$ such that there is an $O(n)$-bundle automorphism $\phi : W|_U \to W'|_{U'}$.

Phrased in these terms, the distinction that Weyl articulates as the a priori versus a posteriori properties of space starts to sound like a rather more familiar distinction in the contemporary literature on spacetime theories: the Anderson–Friedman distinction between *absolute* and *dynamical* structures.[61] To formulate this distinction, Friedman first introduces the concept of *d-equivalence*: given a pair of models $\langle M, \Phi_1, \ldots, \Phi_n \rangle$ and $\langle M, \Psi_1, \ldots, \Psi_n \rangle$ of some spacetime theory T, Φ_i and Ψ_i are d-equivalent just in case 'for every $p \in M$, there are neighbourhoods A, B of p, and a transformation $h : A \to B$, such that $\Psi_i = h\Phi_i$ on $A \cap B$.'[62] Then,

> A geometrical object Φ_i is an *absolute object* of a space-time theory T just in case for any two models $\langle M, \Phi_1, \ldots, \Phi_n \rangle$ and $\langle M, \Psi_1, \ldots, \Psi_n \rangle$ of T, Φ_i and Ψ_i are d-equivalent.[63]

As is well-known, general relativity admits no absolute objects in this sense. But there is a natural way of modifying Friedman's definition of d-equivalence to apply to theories with internal bundles. Given a pair of bundles with fields $(\pi : E \to M, \Phi_1, \ldots, \Phi_n)$ and $(\pi : E \to M, \Psi_1, \ldots, \Psi_n)$, let us say that Φ_i and Ψ_i are *b-equivalent* if for every $p \in M$, there are neighbourhoods A, B of p, and a bundle automorphism $h : E|_A \to E|_B$, such that $\Psi_i = h\Phi_i$ on $E|_A \cap E|_B$; we can then define an absolute object as one which is b-equivalent across any pair of models, rather than d-equivalent.

Hence, if we write any Weylian geometry as $(\pi : E \to M, N, e)$ (where N is the nature, i.e. the action of $O(n)$ on each fibre), then it follows from the observation above that for any models $(\pi : E \to M, N', e')$, N and N' are b-equivalent; and hence, that the nature of the metric is an absolute object in (a natural analogue of) Friedman's sense. So this analysis shows that, even in light of the great conceptual changes wrought by general relativity upon our concepts of space, something fixed and absolute remains. As Weyl puts it:

> One sees that the Riemannian viewpoint does not abnegate the existence of an aprioristic element in spatial structure; only the boundary between the a priori and the a posteriori is shifted.[64]

---

[61] Although this terminology is standard, it is not clear to what extent Anderson's concept of absolute object (as described in Anderson, 1967, §4.3) coincides or overlaps with Friedman's (introduced below): for discussion, see Pitts (2006).

[62] Friedman (1983, p. 58).

[63] Friedman (1983, p. 60).

[64] Weyl (1923b, p. 103).

With this in view, let us return once more to the debate over what represents spacetime. We have seen that the nature of the metric may be understood as a form of structure which goes beyond that of the bare manifold, but is nevertheless absolute or a priori; it is in this sense that it represents an intermediate level of structure between the metric and manifold. That said, it will not succeed in doing the kinds of things that Maudlin or Hoefer argued above that space(time) structure should do. If one is given a manifold with a nature, then there is no way to compute the length of a curve from one point to another; one needs a vielbein field in order to be able to do that. So those who take the metric to be spatio-temporal in character will be unlikely to be persuaded that the (mere) nature of the metric represents spacetime.

References

Anderson, J. L. (1967). *Principles of Relativity Physics*. Academic Press, New York.

Baez, J. C. and Shulman, M. (2010). Lectures on *n*-categories and cohomology. In Baez, J. C. and May, J. P., editors, *Towards Higher Categories*, pages 1–68. Springer, New York.

Coleman, R. A. and Korté, H. (2001). Hermann Weyl: Mathematician, Physicist, Philosopher. In Scholz, E., editor, *Hermann Weyl's* Raum-Zeit-Materie *and a General Introduction to His Scientific Work*, pages 198–270. Springer, Basel.

Curiel, E. (2019). On geometric objects, the non-existence of a gravitational stress-energy tensor, and the uniqueness of the Einstein field equation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66:90–102.

Dewar, N. and Weatherall, J. O. (2018). On Gravitational Energy in Newtonian Theories. *Foundations of Physics*, 48(5):558–578.

Earman, J. and Norton, J. (1987). What Price Spacetime Substantivalism? The Hole Story. *The British Journal for the Philosophy of Science*, 38(4):515–525.

Einstein, A. (1921). *Geometrie und Erfahrung*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Folland, G. B. (1970). Weyl manifolds. *Journal of Differential Geometry*, 4(2):145–153.

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

Handschriften und Autographen der ETH-Bibliothek (1995). Hermann Weyl: Verzeichnis der Manuskripte und Korrespondenzen.

Hawkins, T. (2000). *Emergence of the Theory of Lie Groups*. Springer New York, New York, NY.

Helmholtz, H. v. (1995). On the origin and significance of geometrical axioms (1870). In *Science and Culture: Popular and Philosophical Essays*, pages 226–248. University of Chicago Press, Chicago, IL.

Helmholtz, H. v. (2007). On the factual foundations of geometry (1866). In Pesic, P., editor, *Beyond geometry: Classic papers from Riemann to Einstein*, pages 47–52. Dover, Mineola, NY.

Hoefer, C. (1996). The metaphysics of space-time substantivalism. *The Journal of Philosophy*, 93(1):5–27.

Institut d'Estudis Catalans (1997). *Catàleg de publicacions, 1907–1996*. Institut d'Estudis Catalans.

Knox, E. (2017). Physical relativity from a functionalist perspective. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* (forthcoming).

Levi-Civita, T. (1917). Nozione di parallelismo in una variatà qualunque. *Rend. del Circ. Math. di Palermo*, 42.

Levi-Civita, T. (1922). *Qüestions de Mecànica Clàssica i Relativista: Conferències Donades El Gener de 1921*. Col·lecció de Cursos de Física i Matemàtica. Institut d'Estudis Catalans, Barcelona.
Malament, D. B. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. University of Chicago Press, Chicago, IL.
Maudlin, T. (1988). The Essence of Space-Time. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1988:82–91.
Newman, M. H. A. (1957). Hermann Weyl. 1885–1955. *Biographical Memoirs of Fellows of the Royal Society*, 3:305–328.
Pitts, J. B. (2006). Absolute objects and counterexamples: Jones-Geroch dust, Torretti constant curvature, tetrad-spinor, and scalar density. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 37(2):347–371.
Poincaré, H. (1898). On the Foundations of Geometry. *Monist*, 9(1):1–43.
Poincaré, H. (1952). *Science and Hypothesis*. Dover, New York, NY.
Read, J. (2018). Functional Gravitational Energy. *The British Journal for the Philosophy of Science* (forthcoming).
Riemann, B. and Jost, J. (2013). *Über Die Hypothesen, Welche Der Geometrie Zu Grunde Liegen: Historisch Und Mathematisch Kommentiert von Jürgen Jost*. Springer, Berlin.
Riemann, B. and Weyl, H. (1919). *Über Die Hypothesen, Welche Der Geometrie Zu Grunde Liegen*. Springer, Berlin.
Rovelli, C. (1997). Half way through the woods. In Earman, J. and Norton, J. D., editors, *The Cosmos of Science*, pages 180–223. University of Pittsburgh Press, Pittsburgh.
Ryckman, T. (2005). *The Reign of Relativity: Philosophy in Physics 1915-1925*. Oxford University Press, Oxford.
Ryckman, T. A. (1999). Einstein, Cassirer, and General Covariance — Then and Now. *Science in Context*, 12(4):585–619.
Scheibe, E. (1957). Über das Weylsche Raumproblem. *Journal für die reine und angewandte Mathematik*, 197:162–207.
Scheibe, E. (1988). Hermann Weyl and the Nature of Spacetime. In Deppert, W., Hübner, K., Oberschelp, A., and Weidenmann, V., editors, *Exact Sciences and Their Philosophical Foundations*, pages 61–82. Lang, Frankfurt am Main.
Scholz, E. (2001). Weyls Infinitesimalgeometrie, 1917 – 1925. In Scholz, E., editor, *Hermann Weyl's Raum-Zeit-Materie and a General Introduction to His Scientific Work*, volume 30, pages 48–104. Birkhäuser Basel, Basel.
Scholz, E. (2016). The problem of space in the light of relativity: The views of Hermann Weyl and Elie Cartan. In Bioesmat-Martagon, L., editor, *Eléments d'une Biographie de l'espace Géométrique*, pages 255–312. Presses Universitaires de Nancy, Nancy.
Weatherall, J. O. (2016). Fiber bundles, Yang–Mills theory, and general relativity. *Synthese*, 193(8):2389–2425.
Weyl, H. (1918a). Gravitation und Elektrizität. *Preussische Akademie der Wissenschaften (Berlin) Sitzungsberichte. Physikalisch-Mathematike Klasse*, pages 465–480.
Weyl, H. (1918b). Reine Infinitesimalgeometrie. *Mathematische Zeitschrift*, 2(3-4):384–411.
Weyl, H. (1921). *Raum, Zeit, Materie: Vorlesungen über allgemeine Relativitätstheorie*. Springer Berlin Heidelberg, Berlin, Heidelberg, 4th edition.
Weyl, H. (1922a). Das Raumproblem. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 31:205–221.
Weyl, H. (1922b). Die Einzigartigkeit der Pythagoreischen Maßbestimmung. *Mathematische Zeitschrift*, 12(1):114–146.
Weyl, H. (1923a). *Mathematische Analyse des Raumproblems*. Springer, Berlin.
Weyl, H. (1923b). *Raum, Zeit, Materie: Vorlesungen über allgemeine Relativitätstheorie*. Springer, Berlin, 5th edition.
Weyl, H. (2009). *Philosophy of mathematics and natural science*. Princeton University Press, Princeton.

Friedman and Some of His Critics on the Foundations of General Relativity

Ryan Samaroo

Abstract The paper is an examination of Michael Friedman's analysis of the conceptual structure of Einstein's theory of gravitation, with a particular focus on a number of critical reactions to it. Friedman argues that conceptual frameworks in physics are stratified, and that a satisfactory analysis of a framework requires us to recognize the differences in epistemological character of its components. He distinguishes first-level principles that define a framework of empirical investigation from second-level principles that are formulable in that framework. On his account, the theory of Riemannian manifolds and the equivalence principle define the framework of empirical investigation in which Einstein's field equations are an intellectual and empirical possibility. Friedman is a major interpreter of relativity and his view has provoked a number of critical reactions, nearly all of which miss the mark. I aim to free Friedman's analysis of Einsteinian gravitation from a baggage of misconceptions and to defend the notion that physical theories are stratified. But I, too, am a critic and I criticize Friedman's view on several counts, notably his characterization of a constitutive principle and his account of the principle of equivalence's methodological role.

1 Introduction

There is an approach to the foundations of the exact sciences that is characterized by a certain kind of critical conceptual analysis. This ought not to be confused with the method of analysing notions from ordinary language, of the sort associated with early twentieth-century "linguistic philosophy" and found in certain strains of contemporary analytic philosophy. Rather, the kind of analysis in question—one with a long lineage—is the practice of identifying important features of concepts, and by extension conceptual frameworks, by revealing the presuppositions on which

R. Samaroo (✉)
Somerville College, Oxford, UK
e-mail: ryan.samaroo@philosophy.ox.ac.uk

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_7

their use depends.[1] In the foundations of physics, this analysis is beholden to the body of theory and practice in which concepts are situated and in which they are interconnected with other concepts, both physical and mathematical. A main objective of such an analysis, therefore, is the identification and explication of these connections. While this kind of critical conceptual analysis has several aspects, one of its aims is to reveal what principles are needed for objects of knowledge to be objects of knowledge; in this regard, it is concerned with conditions of possibility, comprehensibility, and meaning.

This kind of analysis was integral to the logical empiricists' approach to the analysis of scientific knowledge. They held that a satisfactory analysis of a theory in the exact sciences should reveal the differences in methodological character of that theory's components. And in this way they defended the notion that our theoretical knowledge is stratified. The idea of such a stratification was criticized by Quine, who argued that there is no reason of principle for distinguishing between the components of our theories. Michael Friedman's approach to the analysis of physical theories is part of a tradition that aims to rehabilitate this aspect of the logical empiricists' account. He defends the stratification of our theoretical knowledge, arguing that the conceptual structures of Newton's and Einstein's theories of gravitation exhibit just the sort of stratification that Quine rejected. In overview, he draws a distinction between a first level of principles that define a framework of empirical investigation and a second level that are made possible by the former. He argues that the theory of Riemannian manifolds and the equivalence principle constitute the framework of investigation in which Einstein's field equations are an intellectual and empirical possibility.

The view at issue here is not the one Friedman defended in *Foundations of Space-Time Theories* (1983), but the view that is found in several works spanning roughly the past twenty-five years, notably "Philosophical Naturalism", his Presidential Address to the American Philosophical Association (1997), *Dynamics of Reason*, his Kant Lectures at Stanford University (1999), and *Synthetic History Reconsidered* (2010). Friedman's approach is a significant contribution to the foundations of physics and the theory of theories. A number of reactions to it have been gathered in *Discourse on a New Method* (Domski & Dickson, 2010) and others can be found. Most of the contributions to this collection pay homage to Friedman. Very little of this work addresses his proposal directly. I will consider a number of challenges to Friedman's approach and especially its application to the analysis of Einsteinian gravitation. Some of these challenges have been ruminating in the foundations of physics for years, but have not been properly articulated and defended. Others are implicit in work that is focused on other goals. A few have not been raised at all.

In what follows, I will outline, in §2, Friedman's approach to the analysis of physical theories, followed by its application, in §3, to Einstein's theory of gravitation. In §4, I will develop and reply to several challenges to Friedman's analysis. I will argue that nearly all of these miss the mark. Through the analysis of

[1] I owe this way of expressing the basic idea of conceptual analysis to Demopoulos (2000, p. 220).

these challenges, I aim to free Friedman's analysis of Einsteinian gravitation from a baggage of misconceptions. But I, too, am a critic and I criticize Friedman's view on several counts. I challenge his account of a constitutive principle and also that of the principle of equivalence. For all that, I defend the notion that physical theories are stratified, and so defend a position in the vicinity of Friedman's.

2 Friedman's Approach to the Analysis of Theories

It is worth situating Friedman's approach to the analysis of physical theories in a broader tradition in the theory of theories. The theory of theories is that part of the philosophy of science that is concerned with the nature of our theoretical knowledge. It is concerned, in particular, with the epistemic status of the *principles* that empirical theories comprise. It asks the following questions: what is the character of these principles? Are the conceptual frameworks that they generate entirely determined by empirical evidence or do they reflect extra-empirical considerations and stipulations? Do all the principles stand on the same footing or do some have a special status and, if so, what is their status? What is the relation of these conceptual frameworks to the world of experience? These questions arise because we have various empirical theories that are well justified. The theory of theories examines, in short, the *basis* for their justification. Furthermore, by clarifying the structure of theories, the theory of theories aims to improve our understanding of the limits of our knowledge of the world, and we acquire a standpoint from which we can better evaluate claims about reality that are consistent with these theories.

This tradition in the theory of theories has its origin in the work of the logical empiricists, and notably in that of Carnap. Carnap took issue with traditional empiricism's claim that *all* knowledge is based on experience. He saw that logic and certain mathematical theories—the latter even in their applications—are not empirically constrained. For example, the statement "$2 + 2 = 4$" is not subject to empirical confirmation or infirmation (Carnap 1963, §10, p. 64).[2] In this way he sought to show that empiricism holds only for empirical principles and not for logical and (certain) mathematical ones.

This view is encapsulated in the thesis that *certain applied mathematical theories are nonfactual*. Demopoulos (2013, Chap. 2) has called it "Carnap's thesis". Carnap's strategy for establishing the thesis rests on his account of analyticity: he held that any statement that is analytic is as nonfactual as a simple tautology.[3] And certainly on one reading of Carnap's account, analyticity can be understood as truth

[2]Carnap was of course well aware that mathematical theories such as geometrical theories—in their applications—have factual content, whereas arithmetic—in its application—does not. He was not concerned, therefore, to distinguish *pure* from *applied* mathematical theories, but rather *applied arithmetic* from *applied geometry*.

[3]Founding analyticity on tautology is, evidently, a Wittgensteinian move: if a sentence expresses a genuine proposition, i.e., is informative, then it partitions states of affairs into those that obtain and

in virtue of meaning, since what is true in virtue of meaning is not informative, that is, is nonfactual.[4]

However we are to establish the thesis that logic and certain mathematical theories are nonfactual, Carnap and the logical empiricists were concerned with what is epistemologically distinctive about these parts of our total body of knowledge. They were concerned to show that these parts of our knowledge have different criteria of truth. Principles tied to the observable, whether directly or indirectly, are "answerable" to experience, in the sense that they are empirically constrained, and the account of their truth, however understood, follows from that. By contrast, the principles of logic and mathematics are not empirically constrained. Their truth rests on different criteria, and, for Carnap at least, this was understood along the lines of Hilbert's proposal that the truth of the axioms of a mathematical theory amounts to nothing more than their consistency.

The analytic-synthetic distinction, though originally drawn in general theory of knowledge, as part of a critique of traditional empiricism, was held by Carnap to be indispensable to the analysis of science. He held that the analysis of the language of science should distinguish between the principles comprising an empirical theory according to their criteria of truth, and in the same measure show how they are integrated into a whole.

The analytic-synthetic distinction was criticized by W. V. Quine, notably in "Two Dogmas of Empiricism" (1951) and "Carnap and Logical Truth" (1960). Quine represented scientific knowledge as a web of belief in which no satisfactory analytic-synthetic distinction can be drawn. In the absence of a suitably broad notion of analyticity, no statements deserve to be singled out as being true in virtue of their meanings or as having any other measure of necessity, apriority or epistemic security. Quine acknowledged that certain stipulations like definitions are undoubtedly analytic, but that we can have no assurance that the principles of mathematics are epistemologically distinguished from physical principles just because they have been stipulated to be analytic. The arbitrariness that attaches to any such stipulation led him to reject the analytic-synthetic distinction. For Quine, all that remains of analytic truth is the centrality of certain statements to the web of belief.

This view, while motivated by a particular understanding of Carnap's and the logical empiricists' approaches to the analysis of theories, led Quine to the far more general view that no distinctions of kind can be drawn among the statements comprising our web of belief. There is no distinction of kind between

---

those that fail to obtain. Tautologies and contradictions do not effect such a partition. Therefore, tautologies and contradictions are not genuine propositions.

[4]Other explications of analyticity include truth in virtue of definition or truth in virtue of convention. Carnap himself referenced Wittgenstein's *Tractatus* in his account of analyticity, and for this reason Demopoulos (2013) referred to this strategy for establishing the nonfactuality of certain mathematical theories as the "Tractarian strategy". But Demopoulos proposes another strategy for establishing Carnap's thesis that he calls the "Einsteinian strategy". This strategy has no precedent in Carnap's writing and it turns on Frege's notion of a criterion of identity.

mathematical and physical principles, and no distinction between these principles and philosophical principles. These principles are all just various strands in the web of belief. Quine called this view "naturalism".

From the view that there are no distinctions of kind between the strands of the web of belief, Quine was led to sketch an alternative account of theories in the final section of "Two Dogmas". This sketch rests on two main ideas. The first is that theories are integrated wholes that are confirmed or infirmed as wholes. This is an appropriation and extension of Duhem's (1962) observation about physical theories. The second is the idea that in the event that the conclusion of a derivation conflicts with experience, there is nothing that prevents us from revising the principles, even the logical and mathematical ones, that figured in the derivation. From this, Quine held that all principles in the web of belief have to some extent an empirical aspect. As we will see, Friedman takes a stance with respect to both of these ideas in his analysis of Einsteinian gravitation.

Now Quine's criticism of the analytic-synthetic distinction was problematic from the start, and many criticisms of it have been raised. There is the classic criticism that Quine's view amounts to scepticism about meaning (Grice and Strawson 1956). It is also questionable whether Quine ever understood Carnap's thesis as a claim that is detachable from some form of conventionalism (Demopoulos 2013, p. 32, n. 6). The most significant criticism is that Quine's critique and sketch of an alternative account of theories fail to draw the factual–nonfactual distinction (Demopoulos 2013, pp. 43–45). I will elaborate on this further on, and with particular regard to Friedman's account. But whatever one's view of the success of Quine's account, it remains that many, if not most, post-positivist philosophers sided with him.

Friedman's view (e.g., 1997, 2001, 2010) is set against Quine's naturalism and his account of the structure of theories that follows from it. Friedman sees in the conceptual structures of Newtonian and Einsteinian gravitation a clear basis for correcting Quine. These theories show that there are differences between the components of our frameworks of physical knowledge, and furthermore that these components are stratified. To anticipate what is to come, Friedman replaces the analytic-synthetic distinction with a distinction between what he calls "constitutive principles" and "properly empirical claims". I will give a brief overview of this account of theories.

On Friedman's account, the analysis of physical knowledge has three levels of enquiry. The *first level* is comprised of constitutive principles that are epistemologically distinguished by the fact that they define a space of intellectual and empirical possibilities, and so determine a framework of investigation. They articulate a framework of theoretical concepts and their physical interpretations. Of these principles, Friedman calls "mathematical principles" those that define a space of mathematical possibilities and that allow certain kinds of physical theories to be developed. They supply a formal background or language that makes it possible to articulate a theory's basic concepts and that makes particular kinds of applications possible. We find, for example, the calculus, linear algebra, and Riemann's theory of manifolds. But there are other constitutive principles that have a more complex character: these "coordinating principles" interpret the concepts that are necessary for physics as we understand it. They express mathematically formulated criteria by which concepts such as force, mass, motion, electric field, magnetic field, space, and time may be applied. In this way the coordinating principles define and articulate our epistemic relation with the world, they fix an interpretation of the world; the mathematical principles, as part of the formal background or language, are auxiliaries or prerequisites to this.[5] Consider what is perhaps the simplest example of a coordinating principle, namely the principle of free mobility that controls the application of Euclidean geometry. This is the principle according to which rigid body may undergo arbitrary continuous motions without change of shape or dimension. Euclidean geometry, which can of course be understood uniquely as an axiomatic system, becomes a theory of physical geometry when it is supplemented with the principle of free mobility, which underlies our ability to perform the Euclidean constructions.

The *second level* is comprised of empirical hypotheses that are formulable within the framework constituted by the first-level principles. For example, in his analysis of Newtonian gravitation, Friedman identifies Euclidean geometry, the calculus, and the laws of motion as constitutive principles of the framework of empirical investigation in which Newton's deduction of the law of universal gravitation, an empirical hypothesis, from the phenomena is an intellectual and empirical possibility.

Friedman also identifies a *third level* comprised of distinctly philosophical or meta-theoretical principles that underlie and motivate discussions of the framework-defining principles, and so the transition from one theory to another. In fact, we find running through Friedman's work a thesis about the nature of revolutionary theory change that I have called "Friedman's thesis"; see Samaroo (2015) for an examination.

With this account, Friedman's principal goal is to restore a proper understanding of the stratification of our conceptual frameworks in physics. His account stands in sharp contrast with Quine's "naturalism" and his related account of theories, according to which there are no differences of methodological principle between the strands comprising the web of belief. This, for Friedman, is the true failure of Quine's account.

3 Friedman's Analysis of Einsteinian Gravitation

Friedman brings this approach to the analysis of physical theories to bear on Einsteinian gravitation. He regards Riemann's theory of manifolds and the equivalence principle as constitutive presuppositions of Einstein's field equations, a properly empirical hypothesis. The former define the framework of empirical investigation in

[5]See Samaroo (2015, p. 130) for further details on the notion of a coordinating principle, with reference to the contributions of Reichenbach and Carnap.

which the latter are an intellectual and empirical possibility. We find this view, for example, in *Dynamics of Reason* (2001):

> [T]he three advances together comprising Einstein's revolutionary theory should not be viewed as symmetrically functioning elements of a larger conjunction: the first two [Riemann's theory of manifolds and the equivalence principle] function rather as necessary parts of the language or conceptual framework within which the third [the field equations] makes both mathematical and empirical sense. (Friedman 2001, p. 39)

To defend these claims, Friedman (2001, 2010) recalls Einstein's argument from the special theory of relativity to the theory of gravitation, stressing the constitutive function of Riemann's theory and the equivalence principle. This account is as follows.

Having shown, in 1905, that simultaneity is not absolute but relative, and having derived the Lorentz transformations from a criterion involving emitted and reflected light signals, Einstein realized that his special theory of relativity clashed with Newtonian gravitation: the latter's hypothesis that there is an instantaneous action-at-a-distance between everybody in the universe is incompatible with the postulate that nothing propagates faster than the speed of light. He realized that a new theory of gravitation was needed to remove the conflict.

The new theory had its origin in Einstein's insight of 1907 into the nature of gravitation. This is the insight, roughly speaking, that bodies in free fall do not "feel" their own weight. Einstein formalized this insight in the principle that we now know as "the equivalence principle". This principle motivates a critical analysis of the inertial frame concept peculiar to special relativity, which we might call "the 1905 inertial frame concept". The inertial frame in question is a frame in uniform rectilinear motion in which the outcomes of all mechanical and electrodynamical experiments are the same.[6]

There are several versions of the equivalence principle. Some are formulated in the context of theory development, others in the context of the completed gravitation theory and exploiting its expressive resources. My focus will be solely on those versions formulated in the context of theory development. Among these, there is a further distinction to be drawn between "gravity-producing" versions of the principle, on the one hand, and "transforming-away" versions, on the other. Both of these can be found in Einstein's own accounts of his theory and its development.[7]

The gravity-producing version is the claim that it is impossible to distinguish between a homogeneous gravitational field and a uniformly accelerated frame. Einstein preferred the gravity-producing version, since true gravitational fields

[6] As would be discovered later in the twentieth century, this is true not only of mechanical and electrodynamical experiments but of all non-gravitational experiments.

[7] Versions of the gravity-producing principle can be found in Einstein's "On the Relativity Principle and the Conclusions Drawn from It" (1907, p. 454), "On the Influence of Gravitation on the Propagation of Light" (1911, pp. 898–899), and in the review article "The Foundation of the General Theory of Relativity" (1916, pp. 772–773). He expressed it as a transforming-away principle in his Princeton Lectures (1922, pp. 67–68). For further details on Einstein's understanding of the equivalence principle, see Norton (1985).

cannot be "transformed away" by free fall. But it is the transforming-away version that is ultimately more important. The transforming-away version, which is an interpretive extrapolation from the principle of the universality of free fall, is the hypothesis that the outcomes of all local non-gravitational experiments are the same as would be obtained in a locally freely falling frame. (Hereafter when I refer to "the equivalence principle" it is to this principle that I am referring.) And what it establishes is that a freely falling frame is locally indistinguishable from a 1905 inertial frame.

We might call the new inertial frame concept that emerges from this analysis "the 1907 inertial frame concept". It is in several respects the cornerstone of Einstein's theory of gravitation. Therefore, the equivalence principle motivates a new inertial frame concept and, with it, a new framework of empirical investigation, one in which the Newtonian and special-relativistic distinction between inertial and non-inertial frames is replaced with a distinction between freely falling and non-freely falling frames. In this framework, Einstein could explore the significance of freely falling trajectories. For all these reasons, Friedman claims that the equivalence principle is a constitutive principle:

> Einstein's field equations describe the variations in curvature of space-time geometry as a function of the distribution of mass and energy. Such a variably curved space-time structure would have no empirical meaning or application, however, if we had not first singled out some empirically given phenomena as counterparts of its fundamental geometrical notions—here the notion of geodesic or straightest possible path. The principle of equivalence does precisely this, however, and without this principle the intricate space-time geometry described by Einstein's field equations would not even be empirically false, but rather an empty mathematical formalism with no empirical application at all. (Friedman 2001, pp. 38–39)

With the 1907 inertial frame concept established, it is worth recalling how Einstein interpreted it in such a way as to make it the basis for his geometrical account of gravitation. The special theory presupposes the mathematical framework of an affine space equipped with a Minkowski metric, and the trajectories of inertially moving particles and light rays are geodesics with respect to that metric. In the special-relativistic framework, gravity is a force that pulls bodies off their rectilinear trajectories. But Einstein had the insight that free fall trajectories might be represented by the geodesics of a variably curved geometry, one determined by the distribution of mass-energy in the universe. This is encapsulated in the geodesic principle, according to which free, massive test-particles traverse time-like geodesics.[8] There were a number of heuristics—all of which falling short of what they needed to establish—that led Einstein to this insight, though Einstein claimed that the "rotating disks" thought experiment was influential.

[8]It is important to note that in this context—the context of theory development—the "geodesic principle" refers to Einstein's insight that the trajectories of freely falling particles might be reinterpreted as geodesics in some yet-to-be-developed theory. But in the context of the completed gravitation theory, there are derivations of the geodesic principle from the field equations.

With the notion that a non-Euclidean and moreover variably curved geometry might be used to represent the trajectories of freely falling bodies, Einstein turned to his friend Marcel Grossmann for assistance. Grossmann introduced Einstein to Riemann's theory of manifolds, which provided the mathematical framework in which the insight summarized in the geodesic principle might be expressed. For this reason, Friedman claims that Riemann's theory of manifolds is a constitutive presupposition of the metrical conception of gravitation that the equivalence principle motivates:

> Without the Riemannian theory of manifolds ... the space-time structure of general relativity is not even logically possible, and so, a fortiori, it is empirically impossible as well. (Friedman, 2001, p. 84)

Together, Friedman claims, Riemann's theory of manifolds and the equivalence principle, are constitutive of the framework of empirical investigation in which Einstein's field equations, a properly empirical hypothesis, are an intellectual and empirical possibility. With these two constitutive principles, we gain a conceptual framework in which it is conceivable that a yet-unknown source-term representing a mass-energy distribution could be related to a yet-unknown geometric object representing chronogeometry.

With this account, Friedman aims to show that, far from there being no distinctions of kind between the components of a framework, the distinctions are in fact significant. Friedman also aims to show that the Quinean notion that any component of a theoretical framework can be revised is baseless: in the case of Einsteinian gravitation, the theory of Riemannian manifolds and the equivalence principle are conditions without which the field equations are not even conceivable.

4 Challenges and Replies

In what follows, I will develop several challenges to Friedman's program. Some of these have been ruminating in the foundations of physics for years, others are raised implicitly in work with other goals, and some have not been raised at all. None of them has been considered carefully in connection with Friedman's view.

4.1 Many Ways to Parse a Theory

Don Howard (2004, 2010) has suggested that there are many ways to parse a theory, and therefore if what is constitutive is relative to a particular parsing, then Friedman's distinction between first-level and second-level principles is arbitrary. By "parsing," it seems to be meant that there are many ways to formulate a theory or to resolve it into its component parts.

Howard (2010, p. 349) suggests that we look to the reconstruction of Einsteinian gravitation due to Ehlers, Pirani, and Schild (1972). On the EPS reconstruction—itself an elaboration of the sketch of Weyl (1918, 1921)—the paths of free particles and light rays are taken as primitive. They define, respectively, projective and conformal structures, and these determine the theory's Lorentzian geometry (up to a scale factor). Howard appeals to the EPS reconstruction to argue that there are ways of formulating Einstein's theory that do not appeal to the equivalence principle, and therefore the equivalence principle cannot be said to be constitutive.

While it is true that there are various ways to formulate a theory or resolve it into its component parts, this challenge is based on a misunderstanding of Friedman's view. Friedman aims to identify those principles that make the field equations an intellectual and empirical possibility, and the principles in question reside in the *context of theory development*. He is concerned with the principles that define the framework of empirical investigation in which a relation such as that expressed in the field equations is conceivable. So, to return to Howard's example, the EPS approach does not define the framework of empirical investigation: it resides in the *context of the completed gravitation theory*. It is a reconstruction that is possible only once we have the completed theory in hand. In this respect, therefore, Friedman's distinction between first-level and second-level principles is not arbitrary, though it is problematic in other respects.

4.2 The Equivalence Principle Is Unnecessary for Developing the Field Equations

The second set of challenges is intended to show that the equivalence principle is unnecessary for the field equations to be an intellectual and empirical possibility, and therefore that it cannot be regarded as a constitutive principle. The challenges rest on the following counterfactual: if Einstein had not developed his field equations in 1915, particle physicists would have 20 years later and without the help of the equivalence principle.

The conjecture rests on the work of numerous twentieth-century and also contemporary particle physicists, who appeal to the massless spin-2, and to a lesser extent the massive spin-0 and spin-2, theories of gravity.[9] These theories assume the framework of relativistic field theory and a graviton field, and from these and other assumptions versions and relatives of Einstein's field equations can be recovered. Massless spin-2 gravity recovers Einstein's field equations in their source-free linearized form. The equivalence principle is satisfied; it becomes a theorem, a consequence or feature of the field equations, rather than a foundational principle. The theory might not be a rival to Einstein's theory itself, but to Einstein's theory with some additional assumptions. The massive spin-0 theory gives a single

[9] See Pitts (2016a, b, 2018) for a list of the original research papers.

equation which is not part of, or logically compatible, with Einstein's equations. The equivalence principle is violated. Furthermore, the theory does not "bend light", so it has been empirically refuted since 1919. It could not be intended as a rival to Einstein's theory in its own right. The massive spin-2 equations are all different from Einstein's equations, albeit in subtle ways. Here, too, the equivalence principle is violated.[10]

These theories suggest two main challenges to Friedman's analysis. The first and most trenchant is implicit in the particle physics approaches to gravitation theory and in the work of Pitts (2016a, b, 2018). This is the view that the equivalence principle is eliminable, and therefore unnecessary for the development of Einsteinian gravitation.

There are two objections to this "eliminativist" view. First, among the alternative theories, only the massless spin-2 theory recovers precisely Einstein's equations, and then only in their source-free linearized form. Second, the equivalence principle is, so far as tests reveal, exceptionless. Therefore, the massive spin-0 and spin-2 theories must, at a minimum, bring something to our understanding of gravitation that outweighs the cost. It is also worth noting that, although it is true that there are multiple paths to (at most) versions and relatives of Einstein's equations, there is a feature of gravitation—the identity of freely falling frames and Lorentz frames—that the equivalence principle singles out. This feature is integral to our understanding of gravitation and the principle not only singles it out but ties it to a number of other concepts. For these reasons, the alternative theories of gravity can hardly be said to support a successful eliminativist account since none of them allows us to recover the full Einstein field equations, which are founded on the principle.

In another challenge directed explicitly at Friedman's account, Pitts (2018, Sect. 3) argues that the equivalence principle is not a constitutive principle, in the sense that it is unnecessary for coordinating the empirical content of Einsteinian gravitation with the field equations. He claims that, while the equivalence principle can fulfil this coordinating role, the principle is unnecessarily strong and some weaker coordinating principle suffices. Pitts bases this view about *Einstein's theory* on the fact that a *massive spin-2 theory* is expected to have nearly the same empirical content as Einstein's theory (when the graviton mass term is sufficiently small).

Pitts' reasoning seems to run as follows: since the equivalence principle is false in massive spin-2 theories, it cannot play a coordinating role. What, then, effects the coordination? Pitts (2018, p. 151) writes: "The coordination gets done ... not by Friedman's principle of equivalence ... Rather, it is done by the field equations ..."[11] Pitts holds that the field equations "themselves" effect the coordination and

[10]In these theories, immersion in a homogeneous gravitational field and uniform acceleration are not identical in their effects. The difference between gravitational effects and inertial effects is observable only in experiments sensitive to the graviton mass term in the gravitational field equation, that is, only if one looks carefully enough to observe the influence of the mass term on inertial effects. See Pitts (2016b, p. 82) for details.

[11]In this, he echoes the remarks of Freund et al. (1969, pp. 861–862) on their massive spin-2 theory.

not an “additional principle” (Pitts, 2018, p. 151). From this view of the coordination of a spin-2 theory with its empirical correlates, Pitts concludes that, similarly, the empirical content of Einstein’s field equations resides in the equations themselves. Therefore, the equivalence principle is not a (coordinating) constitutive principle.

There are several objections to this line of argument. First, it is an odd to argue that Einstein’s theory does not need the equivalence principle as a coordinating principle on the basis of claims about massive spin-2 theories, even if they are found to have nearly the same empirical content in the appropriate limit. The theories in question, though perhaps matching in the appropriate limit, have very different corresponding physical interpretations. Second, even if the geometrical interpretation that we associate with Einstein’s theory has no place in a massive spin-2 theory, the latter still needs some principles to coordinate the basic theoretical concepts that figure in the equations with their empirical correlates. Third, as I will argue in further detail below, it is not the equivalence principle at all that coordinates the empirical content of Einstein’s theory with its basic geometrical notions: it is the geodesic principle that does that. The equivalence principle and the geodesic principle are separate components of the framework of gravitation theory.

4.3 Only Coordinating Principles Are Constitutive

The following challenge is defended in Samaroo (2015). In this and the next section, I develop and refine a few main points.

I have argued that Friedman’s account of a constitutive principle is too broad, and that only coordinating principles should be regarded as constitutive. Friedman’s inclusion of both mathematical principles and coordinating principles in the category of constitutive principles is intended to counter Quine’s contention that the mathematics involved in formulating a theory is just another strand in the web of belief. Friedman argues that this view of the role of mathematics in physics fails to account for the way in which mathematics makes certain kinds of physical theories intellectual possibilities; it also fails to account for the way in which mathematics provides some of the concepts required for formulating a theory and for deriving predictions. I agree with Friedman about this, but there are good reasons for regarding only coordinating principles as constitutive.

The first is that including mathematical principles in a theory’s constitutive component opens the notion of a constitutive principle to trivialization. One might argue that what is constitutive is *relative* to some particular formulation of a theory, and since what is constitutive in one formulation is not constitutive in another, the notion of a constitutive principle is undermined. By taking only coordinating principles as constitutive, we can agree about the principles that interpret the basic theoretical concepts of a given theory, even if that theory admits of an alternative formulation. Consider Newtonian mechanics. The theory admits of various formulations, some of which, e.g., those peculiar to analytic mechanics, rest on radically different mathematical frameworks from the one that Newton presupposed. But however the theory is formulated, Newtonian mechanics is the theory whose basic structure is constituted by the laws of motion. (I will consider the situation in Einsteinian gravitation at the end of this section.)

The second reason is that including mathematical principles in a theory's constitutive component lends support to a main feature of Quine's account of theories, namely "confirmational holism". A Quinean might argue that if the mathematics involved in the formulation of a theory is included in its constitutive component, then the mathematics is confirmed or infirmed along with the rest of the theory. Friedman argues against Quine that constitutive principles are not confirmed in the same way as the empirical hypotheses whose formulation they permit: they are principles without which empirical hypotheses would make neither mathematical nor empirical sense, and without which no test would be possible.[12] The principles that truly establish Friedman's argument against Quine, however, are not the mathematical principles, which, on their own, are subject to neither empirical confirmation nor infirmation, but the coordinating principles that interpret theoretical concepts and control the application of the mathematics. Therefore, distinguishing the mathematical principles from the coordinating principles strengthens the case against Quine.

The third is that including both mathematical principles and coordinating principles in the category a theory's constitutive principles does not draw the distinction that should be drawn between the theory's *factual* and *nonfactual* components, between those components of our theories that are and are not empirically constrained. Taking only coordinating principles to be constitutive allows us to distinguish clearly between those principles that define and articulate our epistemic relation with the world and those that are formal auxiliaries to that. My proposed limitation to the account of a constitutive principle is in no way intended to diminish the role of mathematical principles in the articulation and application of physical theories, nor is it to suggest that they are unnecessary, only to clarify that mathematical and coordinating principles have different criteria of truth. My proposal benefits the account of the stratification of theoretical knowledge and allows for a still stronger criticism of Quine's account to be given.[13]

Now, in reply to these three lines of criticism, one might argue for another account of the stratification of physical theories, for example, Darrigol's "modular" account (2014, forthcoming). Darrigol develops a new account of the relativized a priori, one founded not on constitutive principles but on "comprehensibility conditions". He claims that this account resolves some of the difficulties with Friedman's account, and that it offers a more natural and nuanced account of the

[12]Schematically, the argument is as follows: if Quine's account of theories is successful, then any component, whether mathematical, coordinating or properly empirical, of our total theory is revisable. Some components of our total theory are not revisable in the way Quine would have it because they have a constitutive function. Therefore, Quine's account of theories is unsuccessful.

[13]In several respects, I am arguing for an account of a constitutive principle that is closer to Reichenbach's (1928) account of a coordinative definition, though without any commitment to his view that coordinative definitions are arbitrary.

application, development, and comparison of theories in a given domain. Darrigol's modular account is intricate and a proper exposition is beyond the scope of this article; see his (forthcoming) for a detailed presentation and for a comparison with Friedman's account. But Darrigol's account, like Friedman's, restores the idea that our frameworks of physical knowledge are stratified. In this regard, it certainly is a counter to Quine's account of theories, but it also does not, any more than Friedman's, distinguish between a theory's *factual* and *nonfactual* components. Neither nonfactuality nor relative apriority are properties Darrigol aims to single out.

4.4 The Equivalence Principle and Riemann's Theory Are Not Constitutive

Having considered the case for regarding only coordinating principles as constitutive, let us turn to Friedman's accounts of the equivalence principle and Riemann's theory of manifolds. Is the equivalence principle a (coordinating) constitutive principle? Is it a necessary condition for the field equations to be an intellectual and empirical possibility? Does it coordinate the theory's basic physical notions with geometric notions? While the equivalence principle expands our space of intellectual and empirical possibilities—it motivates a new concept: the 1907 inertial frame concept—what should be clear from the above account, in Sect. 3, is that the equivalence principle lacks the *interpretive* function of a coordinating principle. It is an *empirical hypothesis*—at once an inductive generalization from a set of empirical facts and an interpretive extrapolation from them—and it motivates a new constitutive principle: the geodesic principle. The geodesic principle constitutes or interprets the 1907 inertial frame concept by expressing a criterion for its application: if a test-particle falls freely without rotation, then it moves on a geodesic; if not, its motion deviates from a geodesic, in a way that a yet-to-be-developed theory might measure. The principle coordinates a theoretical concept, the 1907 inertial frame, with a geometric notion, a geodesic. The geodesic principle provides a basis for treating the relative accelerations of freely falling particles as a measure of curvature; in this way, it forms the basis for thinking about gravitation as a metrical phenomenon. It defines a new framework of empirical investigation, one that raises the question to which Einstein's field equations are the answer.[14]

It is worth noting that the equivalence principle and the geodesic principle are separate principles. Of course, in the context of the completed theory of gravitation, the principles are closely related. There is a version of the equivalence principle, due to Anderson (1967) and Ehlers (1973), according to which all non-gravitational

[14]The foregoing is a critical analysis of Friedman's account of the equivalence principle. In other work (Samaroo 2020), I have offered a new account of the principle's methodological role. I have argued that it functions as a criterion of identity for freely falling frames and Lorentz frames.

experiments serve to determine the same affine connection in a sufficiently local region of space-time. The affine connection figures in the geodesic equation, and in this way there is a direct relation between the equivalence principle and geodesics. There is also a derivation of the geodesic equation from the equivalence principle; see Weinberg (1972, Chap. 3, Sect. 2). But Friedman's constitutive principles are found within the context of theory development, so no appeal to these results can yet be made and the equivalence and geodesic principles must be treated as separate parts of the conceptual framework of gravitation theory.

What of Friedman's claim that the theory of Riemannian manifolds is a constitutive presupposition of Einstein's reinterpretation of inertial trajectories as geodesics? The theory of Riemannian spaces is evidently not constitutive in the narrower sense I am defending: it is part of the formal background that made it possible for Einstein to realize his insight that is summarized in the geodesic principle. Some coordinating principle is needed to apply the theory, specifically, the theory of *pseudo*-Riemannian spaces. But is the theory constitutive even on Friedman's account? Friedman emphasizes that a key step in Einstein chain of reasoning was taking spaces of variable curvature to be intellectual and empirical possibilities. But we might distinguish between two things: the transition from the conceptual framework of homogeneous spaces to that of variably curved spaces; the transition from the conceptual framework of variably curved spaces to the mathematical framework of pseudo-Riemannian spaces, which can be regarded as a realization of the former.[15] Both transitions are prerequisites for the development of Einsteinian gravitation, but it is the first transition that seems to be constitutive in Friedman's sense.

To my view that the geodesic principle and not the equivalence principle should be regarded as constitutive, some, e.g., Brown (2005, p. 141 and pp. 161–162 and personal communication), have objected that "it is not simply *in the nature* of force-free bodies to move in a fashion consistent with the geodesic principle", and so the geodesic principle has such limited validity that it could hardly fulfil the (coordinating) constitutive function I attribute to it. This claim is based on the fact that tidal forces act on the constituents of freely falling bodies causing them to spin, and so to deviate from geodesic trajectories.

The geodesic behaviour of free particles is evidently an ideal. But this does not mean that, in the limit in which tidal forces are zero, free test-particles do not exhibit geodesic behaviour. The geodesic principle expresses this ideal which in fact is essential: it is the basis for measuring geodesic deviation (in terms of components of expansion, rotation, and shear), and through this, the basis for learning about the sources of the gravitational field. In Einstein's theory this can be measured.

[15] What is at issue here is the conceptual framework of homogeneous spaces that is picked out by the principle of free mobility. This framework for thinking about physical space was a stumbling block to Poincaré, who, it is conjectured, might otherwise have taken some of the same steps as Einstein towards the gravitation theory.

To my view that only coordinating principles should be regarded as constitutive, one might also object that there is no *unique* way of identifying a given theory's coordinating principles. That is, there is no canonical set. And, if this is so, then one might say that constitutive principles lack a measure of necessity that one would want to attribute to them. The principles do not succeed as conditions of the possibility of the empirical meaning of the field equations.[16]

It is certainly true that there are differences in the accounts of the principles that coordinate the Lorentzian metric to physical events and processes. For example, Malament (2012, pp. 120–121) presents a set of three coordinating principles, which he supplements further on with another involving clocks, and still further on with others involving generic matter fields; Schutz (1985, pp. 182–184) presents another set. Malament's minimal set makes use of only point-particles and light rays; Schutz's employs rods and clocks.

The fact that there are various possible coordinations of the basic physical and geometrical notions should not surprise us, but for this reason it might be said that there is no unique set of coordinating principles. But this would be to overlook what is common to the various coordinations found in relativity texts, namely the geodesic principles for point-particles and light rays. Whatever the differences we find between coordinations, these principles at least are necessary for giving empirical significance to the Lorentzian metric. In this way, therefore, we find something close to the desired uniqueness claim.

5 Significance

Where does the foregoing leave us? Quine claimed to reduce analyticity and apriority to the centrality of certain statements to the web of belief. Friedman is unconcerned with analyticity and retains the a priori in a relativized form: constitutive principles are relativized to particular contexts of enquiry, e.g., the Newtonian and Einsteinian ones, but they determine frameworks of empirical investigation and are in this sense "prior" to the empirical hypotheses whose formulation they permit. But in spite of Friedman's work to restore the idea that conceptual frameworks of physics are stratified, his inclusion of mathematical principles in the category of constitutive principles is a step in the direction of Quine's centrality: it undermines the application of the factual–nonfactual distinction to different components of our conceptual frameworks.

I have argued that those principles that define and interpret basic theoretical concepts should be distinguished from the formal prerequisites or auxiliaries that the principles presuppose, and all of these principles and prerequisites, which together constitute frameworks of empirical investigation, should be distinguished from the empirical hypotheses whose formulation they permit. This allows us to better recognize the salient differences in methodological character. In particular,

[16] I thank two audience members in Bern for raising this objection.

separating mathematical auxiliaries, on the one hand, from coordinating principles and empirical hypotheses, on the other, allows us to distinguish the *factual* from the *nonfactual* components of our theoretical frameworks.[17]

Setting aside these challenges to Friedman's approach to the analysis of theories, my analysis also clarifies several things specifically related to the foundations of Einsteinian gravitation. For one thing, the role of the equivalence principle has been examined. Although there are approaches to relatives and variants of Einstein's field equations that do not appeal to the equivalence principle, the "eliminativist" view, implicit in the work of the particle physics tradition and in the work of Pitts, does not succeed. Furthermore, Pitts' (2018) suggestion that the equivalence principle is not a constitutive principle—in the sense that it is unnecessary for coordinating geometric notions with their empirical correlates—is problematic in several respects. In a final line of argument, I presented and developed the view originally defended in Samaroo (2015). I argued that while the equivalence principle motivates the 1907 inertial frame concept, it is the geodesic principle that constitutes this concept by expressing a criterion for its application. This is the principle that allows us to conceive of gravitation as geometrical phenomenon, and that defines the framework of empirical investigation that permits the formulation of Einstein's field equations.

Far from offering an unqualified defence of Friedman's program or his analysis of Einsteinian gravitation, I have argued that we should critically engage Friedman, but carefully and with criticisms that attain the mark. What my view unequivocally shares with Friedman's is its defence of a stratification of our conceptual framework in physics. Like Friedman, I have defended the importance of identifying the epistemological distinctions between parts of our conceptual frameworks and clarifying their criteria of truth and their functions. And though I can envisage further disagreement about my particular approach to stratification and my replies to the challenges, I hope at least to have freed Friedman's analysis from some of the misconceptions that beset it, and in this way to have strengthened the case against Quine.

References

Anderson, J. L. (1967). *Principles of Relativity Physics*. New York: Academic Press.

Brown, H. (2005). *Physical Relativity: Space-time Structure from a Dynamical Perspective*. Oxford: Oxford University Press, 2005.

Carnap, R. (1963). Autobiography. In *The Philosophy of Rudolf Carnap,* edited by P. Schilpp, pp. 3–86. La Salle, Illinois: Open Court.

Darrigol, O. (2014). *Physics and Necessity: Rationalist Pursuits from the Cartesian Past to the Quantum Present*. Oxford: Oxford University Press.

[17]For an altogether different account of the factual–nonfactual distinction that is independent of the notion of centrality, and also that of a constitutive principle, see Demopoulos (2013, Chap. 2) and Samaroo (2020). This account turns on the notion of a criterion of identity and considers its employment in the foundations of space-time theories.

Darrigol, O. (forthcoming). Constitutive Principles versus Comprehensibility Conditions in Post-Kantian Physics. *Synthese*, doi:https://doi.org/10.1007/s11229-018-01948-2.

Demopoulos, W. (2000). On the Origin and Status of our Conception of Number. *Notre Dame Journal of Formal Logic* **41**, pp. 210–26.

Demopoulos, W. (2013). *Logicism and its Philosophical Legacy*. Cambridge: Cambridge University Press.

Domski, M. & Dickson, M. (eds.) (2010). *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science*. Chicago: Open Court.

Duhem, P. (1962). *The Aim and Structure of Physical Theory*. New York: Atheneum.

Ehlers, J. (1973). Survey of General Relativity Theory. In *Relativity, Astrophysics, and Cosmology*, edited by Werner Israel, pp. 1-125. Dordrecht: D. Riedel.

Ehlers, J., Pirani, F., & Schild, A. (1972). The Geometry of Free Fall and Light Propagation. In *General Relativity: Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh, pp. 63-84. Oxford: Clarendon Press.

Einstein, A. (1907). Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. *Jahrbuch der Radioaktivitat und Elektronik* **4**: 411-62. Vol. 2 of *The Collected Works of Albert Einstein*, edited by J. Stachel, pp. 432-88. Princeton: Princeton University Press, 1989.

Einstein, A. (1911). Über den Einfluß der Schwerkraft auf dei Ausbreitung des Lichtes. *Annalen der Physik* **35**: 898-908. Vol. 3 of *The Collected Works of Albert Einstein*, edited by M. Klein et al., pp. 485-97. Princeton: Princeton University Press, 1993.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik* **49**: 769-822. Vol. 6 of *The Collected Works of Albert Einstein*, edited by A. Kox et al., pp. 283-339. Princeton: Princeton University Press, 1996.

Einstein. A. (1922). *The Meaning of Relativity*. Princeton: Princeton University Press.

Freund, P. G. O., Maheshwari, A., & Schonberg, E. (1969). Finite-Range Gravitation. *Astrophysical Journal* **157**, pp. 857-67.

Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton: Princeton University Press.

Friedman, M. (1997). Philosophical Naturalism. *Proceedings and Addresses of the American Philosophical Association* **71**(2), pp. 5-21.

Friedman, M. (2001). *Dynamics of Reason: The 1999 Kant Lectures of Stanford University*. Stanford, CA: CSLI Publications.

Friedman, M. (2010). Synthetic History Reconsidered. In *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science*, edited by M. Domski & M. Dickson, pp. 571-813. Chicago: Open Court.

Grice, H. P. & P. F. Strawson (1956). In Defense of a Dogma. *The Philosophical Review* **65**(2), pp. 141-58.

Howard, D. (2004). Einstein's Philosophy of Science. In *The Stanford Encyclopedia of Philosophy*, edited by E. Zalta, URL = <https://plato.stanford.edu/archives/fall2017/entries/einstein-philscience/>.

Howard, D. (2010). "Let me briefly indicate why I do not find this standpoint natural." Einstein, General Relativity, and the contingent a priori. In *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science*, edited by M. Domski & M. Dickson, (pp. 333-55). Chicago: Open Court.

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

Norton, J. (1985). What was Einstein's Principle of Equivalence? *Studies in the History and Philosophy of Science* **16**: 5-47.

Pitts, B. (2016a). Einstein's Equations for Spin 2 Mass 0 from Noether's Converse Hilbertian Assertion. *Studies in History and Philosophy of Modern Physics* **56**, pp. 60-9.

Pitts, B. (2016b). Space-time philosophy reconstructed *via* massive Nordström scalar gravities? Laws *vs.* geometry, conventionality, and underdetermination. *Studies in History and Philosophy of Modern Physics* **53**, pp. 73-92.

Pitts, B. (2018). Kant, Schlick and Friedman on Space, Time and Gravity in Light of Three Lessons from Particle Physics. *Erkenntnis* **83**, pp. 135–161

Quine, W. V. (1951). Two Dogmas of Empiricism. *The Philosophical Review* **60**, pp. 20–43.
Quine, W. V. (1960). Carnap and Logical Truth. *Synthese* **12**(4), pp. 350–74.
Reichenbach, H. (1928). *The Philosophy of Space and Time*. Translated by M. Reichenbach & J. Freund. New York: Dover, 1958.
Samaroo, R. (2015). Friedman's Thesis. *Studies in History and Philosophy of Modern Physics* **52**, 129–38.
Samaroo, R. (2020). The Principle of Equivalence as a Criterion of Identity. *Synthese* **197**(8), pp. 3481–3505.
Schutz, B. (1985). *A First Course in General Relativity*. Cambridge: Cambridge University Press.
Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. London: Wiley.
Weyl, H. (1918). Reine Infinitesimalgeometrie. *Mathematische Zeitschrift* **2**, pp. 384–411.
Weyl, H. (1921). Zur Infinitesimalgeometrie: Einordnung der projektiven und der konformen Auffasung. *Nachrichten der Königlichen Gesellschaft Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, pp. 99–112. Reprinted in *Space-Time-Matter* and translated by H. Brose. New York: Dover, 1950.

Interpretations of GR as Guidelines for Theory Change

Niels Linnemann

Abstract One hundred years of relativity—but what is next? In this chapter, I explore to what extent GR itself suggests particular (candidate) successor theories—or rather, to what extent the various empirical and conceptual interpretational stances one can take concerning GR do. At a more general level, the chapter thus aims at demonstrating how interpretational questions allow for systematically generating hypotheses about a successor theory to GR, and thus conceiving of the theory change from GR as—at least potentially—a well-guided process.

1 Introduction

One hundred years of relativity—but what is next? In this chapter, I explore to what extent GR itself suggests particular (candidate) successor theories—or rather, to what extent the various empirical and conceptual interpretational stances one can take concerning GR do. At a more general level, the chapter thus aims at demonstrating how interpretational questions allow for systematically generating hypotheses about a successor theory to GR, and thus conceiving of the theory change from GR as—at least potentially—a well-guided process.

In Sect. 2, I first clarify the notion of interpretation by using a threefold distinction by Curiel (2009). I will be concerned with the straightforward empirical interpretation of a theory, that is what Curiel calls concrete interpretation (sometimes also just empirical interpretation in the following), and the high-level conceptual interpretation of a theory, that is what Curiel calls categorial interpretation. In Sect. 3, I consider the rather unusual question in how far empirical interpretation (concrete interpretation) can provide a guideline for a successor theory to GR. I identify two potential strategies for suggesting a direction in theory change from empirical interpretation: (1) On the one hand, we can ask at a general

N. Linnemann (✉)
University of Bremen, Bremen, Germany
e-mail: Niels.Linnemann@uni-bremen.de

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_8

level what other theory-external descriptions of the established empirical content of the theory there are, and whether these should not be incorporated as well. (2) On the other hand, we can wonder whether specific options for increasing the empirical content of GR and the empirical access to it *via using external theories* can encourage the unification of GR with exactly these external theories. In Sect. 4, I deal with the more common case of motivating successor theories to GR from its conceptual interpretations; I present some examples of how different conceptual takes on GR encourage successor theories to it.

2 Empirical Interpretation(s) of a Physical Theory

In dealing with the question of how to make sense of GR's empirical content, Curiel (2009) distinguishes three notions of empirical interpretation:

> **Concrete.** The fixation of a semantics for the formalism, in the sense that the formalism under the semantics expresses the empirical knowledge the framework contains—for example, the fixation of a Tarskian family of models, or, less formally, the contents of a good, comprehensive textbook.
> **Categorial.** The explication of concepts in the theory that the semantics of a concrete interpretation alone does not fix—for example, a demonstration that the theory is deterministic in any of a variety of senses.
> **Metalinguistic.** The explication of the semantics of a concrete interpretation, when the representational nature of the concrete interpretation is itself not understood—for example, the Copenhagen Interpretation of quantum mechanics (p. 46).[1]

Following Curiel, *all* physical theories require a concrete and a categorial interpretation. More precisely: (1) All physical theories have to make contact with the world in some way or the other in order to count as empirical theories. Thus, there should be no dispute about that physical theories need to be interpreted concretely. (2) Whether a categorial interpretation is necessary or not, does not amount to a case-by-case question but rather to a general question of interest independent of the nature of specific physical theories. Examples for categorial questions include whether a theory is deterministic or indeterministic, whether it allows for superluminal propagation or whether it is observer-independent or -dependent. At least from the view of a philosopher of physics, *all* physical theories should be worth being interpreted categorially in some way or another. And in any case, a categorial question is either raised for all/a subset[2] of theories or not at

[1] Admittedly, it is not really clear why this interpretation is best dubbed 'metalinguistic', since—as we will see—it can also be understood as a fine-grained correction or addition to the concrete interpretation.

[2] Certain categorial questions such as whether a theory allows for superluminal propagation require a background framework to be even sensibly formulated in the first place. They thus only make sense for a subset of theories admissible to this overarching framework. See for instance Weatherall (2014) for a discussion of theory-overarching criteria for superluminal propagation.

all. More metaphysically flavoured questions such as whether a spacetime theory is substantivalist or relationist, etc. are arguably categorial as well. We will, however, ignore questions like this in the following as we are interested in empirical senses of interpretation only.

In contrast, only a few physical theories, such as QM[3] require a metalingustic interpretation. That necessity of a metalinguistic interpretation is a case-by-case question should be clear from that Newtonian mechanics, and QM, respectively, provide paradigm examples for theories which do not have and have to be interpreted metalinguistically. As an illustration, consider that in both Newtonian mechanics and QM the Hamiltonian's value is assigned in one way or the other to energy. Whereas in the former, this assignment holds without restriction, this is not at all the case in QM:[4] The Hamiltonian H is assigned to a measurable energy E in the concrete interpretation but then further qualification is needed: The Hamiltonian H can only be assigned to a measurable energy E upon some story of collapse of a generic state into an eigenstate of H, as $H|\psi\rangle = E|\psi\rangle$ iff $|\psi\rangle$ is an eigenstate of H.

(Strictly speaking, QM only has to be given some kind of qualificatory interpretation for which a metalinguistic interpretation might be just one way to do so: The Copenhagen interpretation fleshes out in how far operators are 'observables' and in particular in what sense experiments show determinate results by providing the primitive theory-external notion of collapse—this is an instance of metalinguistic interpretation. As an alternative to this, the many-worlds interpretation arguably provides a non-trivial internal qualification of what happens in measurement through invoking the idea of a virtual, branch-relative equivalent of an objective collapse. Both the Copenhagen as well as the many-worlds 'interpretations' can be seen as qualificatory interpretations on top of the *usual* concrete interpretation of QM. For a detailed explication of the notion of qualificatory interpretation, which is beyond the scope of this short overview on notions of empirical interpretations, see Linnemann, 2019.)

[3]Note that by 'usual', we mean textbook QM. This is not to say that the concrete interpretation of QM cannot drastically vary either. GRW, for instance, is not just a metalinguistic interpretation of QM but at the same time has a different concrete interpretation than standard QM. Within GRW, the characteristic spontaneous localisation process (including an avalanche of localisations induced by a spontaneous localisation) is part of the empirically measurable content.

[4]Cf. Curiel (2009, p. 47):

> The way that Hermitian operators in standard quantum mechanics represent observables is perhaps the canonical example of such a problem: we know they do in some way or other represent observables, and we know how to use them to construct good models of systems that we can use to predict the (probabilistic) outcomes of experiments, but we have no clear understanding at all of the nature of the representational relations between, on the one hand, the operator as part of the formalism and, on the other, the actual values we measure for physical quantities in experiments.

2.1 The Concrete Interpretation in More Detail

The notion of concrete interpretation can be straightforwardly cashed out in the rather narrow positivist notion of partial interpretation (see Suppe 2000):[5]

> For positivists, theories were partially interpreted axiomatic systems TC where the axioms T were the theoretical laws expressed in a theoretical vocabulary V_T; C were correspondence rules that connected T with testable consequences formulated using a separate observational vocabulary V_O. Only V_O sentences were given a direct semantic interpretation (p. S103).

On such a syntactic understanding of theories,[6] the concrete interpretation fixes (and it alone fixes) by assignments the semantics necessary to test the theory for *empirical adequateness*. It is commonly claimed that the syntactic view of theories has by now been superseded by the semantic view of theories[7] for a set of well-known reasons: among other things, the syntactic view (unlike the semantic view) is usually seen to (1) suffer from issues of theory individuation ($T_1 = p \rightarrow q$ and $T_2 = \neg p \vee q$ are logically equivalent but—oddly—different syntactic theories), (2) make an untenable theoretical/observational distinction at the level of language (rather than that of entities), and (3) narrowly require theories to be formulated in first-order logic (see Winther 2016, and Suppe 2000). Recent arguments, however, render the syntactic view as capable of dealing with these challenges (see, for instance, Lutz

---

[5]Note that a sensible partly interpretation (which the concrete interpretation amounts to) does not have to follow mere empirical interests. For instance, Lehmkuhl (2017) exemplarily pleads for a partly (what he calls 'careful') rather than literal interpretation of the Schwarzschild metric field even within a realist context when the metric is invoked in modelling the perihelion motion of Mercury:

> What Einstein really does is to convert the two-body problem Sun-Mercury into a one-body problem, where one body (Mercury) is subject to an external gravitational field. . . . It is the exterior gravitational field of the Sun, not the Sun itself, that is represented by the Schwarzschild metric. And that is enough to predict the perihelion of Mercury: we do not need to know what the Sun is made of or what happens in its interior; all that matters is the exterior gravitational field that Mercury is subject to. Thus, worrying about the singularity at the center of the Schwarzschild metric misses the point: we do not have to interpret the interior part of the Schwarzschild metric literally, at least not in this application (pp. 1210–1211).

So, the Schwarzschild metric's interior structure should not be seen as referring literally (not even approximately) but rather as a placeholder for any kind of interior structure that (1) would be compatible with the exterior metric field of the Schwarzschild solution, and (2) would involve structure which can in the end be straightforwardly interpreted as referring to the sun.

[6]The syntactic view renders a theory as a set of statements T closed under logical implication (subset of statements, called axioms, generates T when requiring logical closure), and correspondence rules between the entities in T and the world (see Winther, 2016, section 2).

[7]The semantic view renders a theory as a set of models. In the context of a physical theory, you can loosely think of each solution to the dynamical equations for a given set of initial conditions as a model. See Winther (2016), in particular section 3.1.2, and references within. A locus classicus is Van Fraassen (1980).

2014), and as basically equivalent to the semantic view (see Halvorson 2013 and Lutz 2017).

Since we are only interested in denoting the empirical content of a theory, we strictly speaking better stay uncommitted to the existence of the non-observable part of the formalism (constructive empiricism). With the syntactic view, we, however, run into the so-called closure problem:[8] given that belief statements are normally closed under logical implication, fixing referents for the empirical (and just for the empirical) on the syntactic view of theories will automatically commit one to that certain theoretical terms refer as well. A statement which mixes observable and theoretical terms such as “A table is a swarm of electrons, protons and neutrons." would, for instance, ground the existence of electrons, protons and neutrons in the world based on that we believe that tables exist in the world. This would run counter the idea that the concrete interpretation, when defined syntactically as above, is *just* an empirical interpretation of the theory. Rather, partial interpretation as defined in the positivist sense above goes along with a certain commitment as to which purely theoretical concepts (such as electrons) refer. Now, as one normally does not want to give up the logical closure of statements of belief, the problem should be evaded by either switching from the syntactic view to the semantic view of theories, in particular by making use of the ‘partial structure approach[9] or by taking a fictionalist stance on theoretical terms on which we can talk about protons, neutrons, electrons, etc. as if they existed in the world without committing that they actually do (see Rochefort-Maranda 2011). In the following, I will implicitly opt for the latter option.

None of the above is to say that the idea of concrete interpretations should or can only be cashed out by correspondence rules (on the syntactic view) or partial structures (on the semantic view):

- Uninterpreted theoretical terms may be linked to already interpreted ones (internal interpretations). Among others, (1) Stein (1994) proposes connecting a formal theory to the empirical by schematising the observer within the theory, taking it to be straightforward how the observer should then itself be linked to the world. (2) de Haro and de Regt (2018) provide examples of how already interpreted elements can be used for an interpretation of uninterpreted elements provided that they stand in a relevant relationship (such as symmetry).
- Correspondence rules may employ an (approximate) correspondence of a theoretical term to that of another theory for which the linkage to the observable is readily established (external-theory interpretation). Typically, external theories *to which the theory in question knowingly reduces* are used for this purpose. The so-called Schwarzschild mass in the Schwarzschild metric, for instance, can be to some degree interpreted by identifying its role as mass within the Newtonian limit theory.

[8] First introduced by Friedman (1982). See also Rochefort-Maranda (2011).

[9] On the partial structure approach, the model is split up into a part which is known, and a part which is not known to refer. This allows for denoting the observable-part to refer and the un-observable to not refer (see Bueno, 1997, section 3).

2.1.1 The Dynamical Nature of the Concrete Interpretation

So far it may have sounded as if the concrete interpretation was easily fixed once and for all for a theory. However, the concrete interpretation of a theory—qua activity—is, in fact, an ongoing affair. It is, for instance, in this sense that we can read the following passage by Carnap (1966) on correspondence rules:

> Of course, physicists always face the danger that they may develop correspondence rules that will be incompatible with each other or with the theoretical laws. As long as such incompatibility does not occur, however, they are free to add new correspondence rules. The procedure is never-ending. There is always the possibility of adding new rules, thereby increasing the amount of interpretation specified for the theoretical terms; but no matter how much this is increased, the interpretation is never final (p. 238).

Similarly, if you just loosely commit to cashing out the notion of concrete interpretation in terms of the contents of a 'good textbook'—the meme Curiel uses to display what he means by concrete interpretation—it should be clear that there is no definitive, final textbook to be expected on a topic like GR anyway.

In the following, I will make use of what I call empirical access to and empirical content of a theory which I define using the theoretical-observational distinction, and the notion of correspondence rules: The *empirical access* to a theory is given by the correspondence rules between theoretical and observational terms. It is increased when correspondence rules between a theoretical vocabulary and the observational vocabulary are added. The *empirical content* of a theory is given by the observational terms which are actually linked up to the theoretical terms by correspondence rules. It is increased when correspondence rules between a theoretical vocabulary and specific observational vocabulary which had neither directly nor indirectly (that is via other observational terms) been linked yet to theoretical terms become established. Increase in empirical content always implies an increase in empirical access but not vice versa.

The problem with the notion of empirical content is that it is highly dependent on what we take the observational vocabulary to be, which itself is, however, a problematic notion. As Carnap (1966) remarked,

> To a philosopher, "observable" has a very narrow meaning. It applies to such properties as "blue," "hard," "hot." These are properties directly perceived by the senses. To the physicist, the word has a much broader meaning. It includes any quantitative magnitude that can be measured in a relatively simple, direct way. A philosopher would not consider a temperature of, perhaps, 80 degrees centigrade, or a weight of $93\frac{1}{2}$ pounds, an observable because there is no direct sensory perception of such magnitudes. To a physicist, both are observables because they can be measured in an extremely simple way. The object to be weighed is placed on a balance scale. The temperature is measured with a thermometer. The physicist would not say that the mass of a molecule, let alone the mass of an electron, is something observable, because here the procedures of measurement are much more complicated and indirect (pp. 225–226).

So, we will mainly focus on the notion of empirical access in the following. But even here one might be sceptical: When adding correspondence rules linking theoretical terms from T to observational terms from O via making recourse to external vocabulary from T', it is important to make sure that we are not just artificially inserting notions from T' in-between notions from T and O. What we have to require, thus, is that these additional paths from notions of T to notions of O via notions of T' feature notions of T' in a non-redundant fashion of one form or the other.

2.1.2 GR's Concrete Interpretation

A minimal concrete interpretation of GR includes the (potential) association of null geodesics with light rays, time-like curves with point particles of positive mass, and time-like geodesics with free point particles of positive mass (see, for instance, the interpretive principles (C1), (C2), and (P1) in (Malament, 2012, pp. 120–121)).

The, arguably, most minimal scheme for empirical access to the metric field in GR, then, builds on what is called the causal-inertial method: Based on a theorem by Weyl (1921) and[10] Ehlers et al. (2012) provide a prescriptive scheme for determining the metric from the movement of light rays (linked to what's known as conformal structure) and freely falling particles (linked to what's known as projective structure) alone: an observer hereby tracks light rays and freely falling particles relative to local radar coordinates which use an arbitrary parameterisation of the observer's world-line as time parameterisation. It is important to stress that the scheme assumes that sufficiently well-defined radar coordinates can be set up in the first place.[11] Once the observer is thus schematised in the theory (cf. Stein 1994), the metric structure can be interpreted internally through the procedure on how the observer measures out the metric structure using just the few elements of the theory already linked to the world (null geodesics, and time-like paths).

The causal-inertial method for accessing the metric field provides an internal interpretation of the metric field based on the minimal concrete interpretation of null geodesics and time-like curves. Similarly, one could adhere to a theorem by Fletcher (2013) for arguing that null geodesics (corresponding to light) as well as time-like curves (corresponding to massive particles) can be used to construct clocks *within the general relativistic model that read out the world-line interval*. Alternatively, one might, however, also interpret the metric structure through correspondence rules directly. The world-line interval, for instance, is linked by brute force to the reading of an ideal clock on the chronometric approach of Synge (1960).

[10]The theorem basically states that "the projective and conformal state of a metric space determine the metric uniquely" (see Coleman and Korte, 1980, Theorem 4.3).

[11]The scheme was subsequently heavily improved; Coleman and Korte (1980), in particular, managed to free the scheme from charges of circularity. See Bell and Korté (2016) for a summary.

It is also worth noting that, in regimes of weak gravitation, empirical content of GR can in a limited sense be fixed through Newtonian and post-Newtonian approximation of GR (external interpretation); as already mentioned before, the so-called Schwarzschild mass, for instance, acts effectively as a Newtonian point mass on observers who are far enough away from the inner region of the Schwarzschild spacetime.

3 Guidelines From GR's Empirical Interpretation

We can use the concrete interpretation of GR in at least two ways to get a glimpse at what kind of theory succeeds GR: (1) We can take the usually attributed empirical content seriously at a general level, that is wonder what other external theory-descriptions of the established empirical content there is, and whether these external theory-descriptions should not be merged with GR's formalism. This straightforwardly suggests a change of the formalism, and thus provides a direct guideline for theory development. (2) We can take the empirical content of the theory seriously at a more specific level: (a) We can consider whether—also under adherence to external theories—elements of GR's formalism should, after all, be empirically interpreted, that is linked to the empirical by some correspondence rule. Call the posit behind this the *principle of maximal concrete interpretation* (PMCI): A theory's potential concrete interpretation should be exploited to a maximum degree, that is one should strive for the maximal (in principle) empirical content associable to a given theoretical formalism even if this may involve taking into account extratheoretical elements into its semantics. (b) We can consider whether—under adherence to external theories—the empirical content of GR becomes accessible in new ways. Call the posit behind this the *principle of maximal empirical access* (PMEA): One should strive for as many (in principle) modes of access possible to the empirical content even if this may involve taking into account extratheoretical elements into its semantics. Positive results for (a) and (b) both *suggest* merging the external theories' formalism adhered to with GR's formalism. More precisely, when extending the standard theory's empirical interpretation through invoking an external theory, both the theory to be interpreted more and the theory invoked for this become to a certain extent entangled at the level of empirical interpretation: their empirical interpretations are not independent of each other anymore but at least one interpretation now also does recourse to the other. Such a strong entanglement at the semantic level then also suggests some sort of merging of the theories at the formal level. I now illustrate these rather abstract strategies (1) and (2). Thereby, I make two natural heuristics for motivating theory change into a particular direction explicit which are both based on quasi-empirical consideration.

3.1 General Approach

GR-matter fields[12] $\phi_1, \ldots, \phi_n$ such as the electromagnetic field-strength tensor F_{ab} are linked to elements in the observable regime which—provided that curvature effects are negligible—are known to be locally more precisely described as quantum fields in flat spacetime than classical fields in curved spacetime. Thus,

- GR should be enhanced to take into account that the fields $\phi_1, \ldots, \phi_n$ are locally better described as quantum fields in flat spacetime than as classical fields (provided that local curvature effects are negligible). This leads to QFT in curved spacetime, which treats quantum field theories in curved but static Lorentzian background geometries; and, as QFT in curved spacetime neglects back-reaction of the field to its now quantum matter content, to semi-classical gravity, which builds around the semi-classical field equations $G_{ab} = 8\pi \langle \hat{T}_{ab} \rangle$, where $\langle \hat{T}_{ab} \rangle$ is the expectation value of an operator-valued energy-momentum tensor (see Wald 1994).
- The field equation links the field g (via the Einstein tensor G_{ab}) to the matter fields $\phi_1, \ldots, \phi_n$ (via the energy-momentum tensor T_{ab}), i.e. $G_{ab} = 8\pi T_{ab}$. The referents of the right-hand side of this equation are—as just stressed—known to be locally modelled as quantum theories in flat spacetime (provided that local curvature effects are negligible). This suggests then that G on the left-hand side, or rather g making up G, has referents that should be formulated as a quantum theory as well, and thus the quantisation of GR as such—at least if the semi-classical account turns out to insufficient.

3.2 Specific Approaches

I present two examples—one from chronometry, the other from black hole thermodynamics—of how the concrete interpretation of GR can be extended under adherence to the (PMEC) or the (PMEA). Only the second of the two examples, however, provides a relevant insight as to where a successor theory to GR should be heading, which shows us that not every increase in empirical content or empirical access provides non-trivial hints.

[12] In the standard formalism, GR is formulated as a theory of fields on a 4-dimensional manifold with a symmetric rank-two tensor, the metric g of Lorentzian signature, and other (tensor) fields $\phi_1, \ldots \phi_m$ (such as the electromagnetic field-strength tensor F_{ab}) 'on top'. The fields $\phi_{i_1}, \ldots \phi_{i_n}$ which contribute to the energy-momentum tensor T are then called matter fields.

3.2.1 Chronometric Interpretation

Empirical access to the world-line interval in GR can be largely extended through the purely interpretative (albeit not necessarily empirically adequate) stipulation that external clocks—not modellable within the GR-framework itself—read out the world-line interval of the path they are travelling along (i.e. fulfil what has become known as the clock hypothesis[13]).[14]

How well this stipulation works, can be checked (for instance) using an operationalist criterion by Perlick (2008). So, whether a clock is a standard clock or not, is tested by tracking the movement of freely falling particles and light in (local) radar coordinates that are induced by the clock under consideration. More precisely, the clock in question is used to provide a parameterisation for the observer's world-line. Using this parameterisation and light signals, the observer can set up radar coordinates: Objects away from the world-line are assigned a radial distance of $R = 1/2(t_2 - t_1)$ and a time $T = 1/2(t_2 + t_1)$ where t_1 is the parameter value on the world-line which corresponds to emission, and t_2 is the parameter value on the world-line which corresponds to detection of the probing light signal. A clock reads out proper time at clock time t_0 (that is, is a standard/ideal clock) if and only if

$$\left.\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\right|_{t=t_0} = \pm\left.\frac{\frac{d^2R'}{dT'^2}}{1-(\frac{dR'}{dT'})^2}\right|_{t=t_0} (*)$$

where (R, T) and (R', T') denote the radar coordinates for two free particles, respectively, sent off at $t = t_0$ into the same direction with differing speed ('+') or into opposite directions with arbitrary speeds (apart from that at least one needs to have a non-zero speed) ('$-$'). In other words, the extent to which a clock approximates a standard clock 'at a point' can be measured through the difference $\Delta = \left\| \left\| \left.\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\right|_{t=t_0} \right\| - \left\| \left.\frac{\frac{d^2R'}{dT'^2}}{1-(\frac{dR'}{dT'})^2}\right|_{t=t_0} \right\| \right\|$ with $\|..\|$ being a suitably chosen norm. For a freely falling standard clock, even $\left.\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\right|_{t=t_0} = 0$, i.e. $\left.\frac{d^2R}{dT^2}\right|_{t=t_0} = 0$. That is, the radar coordinates induced by a freely falling standard clock make the connection coefficients for the radial acceleration equation vanish at $t = t_0$.

Why should, however, a device fulfil $(*)$? That some external element (like an atomic clock) not at all describable in a general relativistic framework can be found to fulfil this condition, might be argued for from some selection process:

[13]See (Maudlin, 2012, chapter 5), and Fletcher (2013).

[14]It is important to note, however, that *external* clocks are not required in *many* spacetimes for gaining chronometric access to the metric field: As a theorem by Fletcher (2013) demonstrates, light clocks internal to GR can be set up to measure the world-line interval of the metric up to arbitrary precision—provided that light can be said to move on null geodesics.

we have simply managed to identify an ‘apparatus’ (such as an atom’s oscillatory states) which suits our needs. The question of what ultimately makes this selection process possible in the first place remains. Now, the best explanation for why the atomic clock can be used to measure out the world-line interval is then simply that there is some sense in which quantum field theory (describing the atomic clock) and GR (describing the world-line interval) go along with each other. The hereby suggested unification is a bit sobering, however: we already know much more straightforwardly—namely from that the matter content adhered to in GR usually have a quantum description in approximately locally flat spacetime regions—that some kind of theory like QFT in curved spacetime/semi-classical gravity is needed (see Sect. 3.1).

(As an alternative to the straightforward chronometric interpretation, one could—in tradition of the dynamical approach to relativity[15] stipulate the strong equivalence principle (SEP)—loosely speaking, the idea that the matter field dynamics in GR is somewhat locally Minkowskian—as a means of establishing operational access to the metric via theory-external clocks:

> ... [the strong equivalence principle] allows us to carry over certain interpretational possibilities from SR. In particular, it allows us to transfer the interpretation of rods and clocks as waywisers of the metric tensor from the special case of the Minkowski metric to the case of a generically curved (but locally Minkowskian) metric, and it allows us to interpret the frames of reference in which the metric is locally Minkowskian as local inertial frames in the sense of ‘inertial frame’ we are wont to use from SR. The local validity of SR allows a ‘trickling up’ of interpretations from SR to GR. I said that this makes the role of the SEP seem interpretational, but we have to be careful not to see its role as ‘merely’ interpretational. The SEP explains why rods and clocks can serve as waywisers of the metric field (Lehmkuhl 2011, p. 26).

The only problem with this is that currently known versions of the SEP are not satisfactorily formulated to this purpose.

3.2.2 Thermodynamic Interpretation

When seen within the narrow context of GR, the so-called black hole thermodynamic laws are statements on the relationship between geometry, and the energy and charges at asymptotic infinity of the spacetimes to which they apply; they are analogous to the standard thermodynamic laws (see Table 1). However, once understood in the context of semi-classical GR (which clearly goes beyond GR proper)—in particular in light of Hawking’s derivation of what is now called Hawking radiation—they are typically interpreted as proper laws of thermodynamics.[16]

[15] See Brown (2005), and, relatedly, Knox (2013).

[16] At least in the physics community. For critical accounts on the status of black hole thermodynamics qua thermodynamics, see Wüthrich (2019) and Dougherty and Callender (forthcoming); for a defence of the orthodox view on black hole thermodynamics as more than an analogy (see Wallace, 2017).

Table 1 Overview of the analogies between standard thermodynamic laws and stationary black hole thermodynamics (largely taken from Kiefer, 2004, p. 202)

	Standard thermodynamics	Black hole thermodynamics
Zeroth law	T constant on a body in thermal equilibrium	κ constant on the horizon of a stationary black hole
First law	$dE = TdS - pdV + \mu dN$	$dM = \frac{\kappa}{8\pi G}dA + \Omega_H dJ + \Phi_H dQ$
Second law	$dS \geq 0$	$dA \geq 0$
Third law	$T = 0$ cannot be reached	$\kappa = 0$ cannot be reached

On the thermodynamic side, T is the temperature, S the entropy, p the pressure, V the volume, μ the chemical potential, and N the particle number of the standard thermodynamic system. On the black hole side of the analogy, κ is the surface gravity, A is the horizon area, Ω_H the angular velocity, J is the angular momentum, ϕ is the electrostatic potential, and Q the electric charge of a black hole

But what keeps us from *interpreting* the surface gravity[17] starring in the black hole thermodynamic laws as thermodynamic temperature already at the level of GR proper? This association is (in principle) falsifiable, and well-motivated from the analogy (never mind that classical black holes do not allow for escape from its internal region, and are thus associated, if at all, with a temperature $T = 0$).[18] I consider now in what senses GR's black hole thermodynamic laws call for a tentative extension of GR's concrete interpretation.

More precisely: GR has a theoretical vocabulary T_{GR}, and thermodynamics that of $T_{\mathrm{Thermodynamics}}$. The theory-overarching observational vocabulary is O. Correspondence rules $C(T_{\mathrm{GR}}, O)$ connect elements from T_{GR} to elements in O. Notably, no correspondence rules on the standard conception links any theoretical term from GR to the "sensation of temperature", $o_{\mathrm{heatsensation}} \in O$. Now, stipulate that the surface gravity $\kappa \in T_{\mathrm{GR}}$ corresponds to the temperature $\in T_{\mathrm{Thermodynamics}}$, that is that there is an inter-theoretical relation "$\kappa = c \cdot T$" $\in C'(T_{\mathrm{GR}}, T_{\mathrm{Thermodynamics}})$, where $C'(T_{\mathrm{GR}}, T_{\mathrm{Thermodynamics}})$ denote correspondence rules between the theoretical vocabulary of GR and that of thermodynamics, and c is a constant. Under this assumption, $\kappa \in T_{\mathrm{GR}}$ could be associated to (1) the point coincidences observed on a thermometer $o_{\mathrm{pointcoincidence}} \in O$, and arguably even to (2) the sensation of heat $o_{\mathrm{heatsensation}} \in O$.

If case (1) is empirically established, *the empirical access linked to GR's empirical content is increased*, as now additional correspondence rules between

[17] Or a multiple thereof. After all, the first law only allows for an association of T with κ and S with A up to a positive constant k, i.e. $T = k\kappa$ and $S = \frac{A}{k}$.

[18] In semi-classical gravity, this association is derived as what is known as Hawking radiation. For the original paper (see Hawking, 1975). It was this finding which made physicists finally believe that black hole thermodynamics is more than an analogy. The point here is that this association would be worth trying *even if we had no hints from a theory beyond GR for its potential validity yet*. For a consequent albeit heavily heterodox plea for why even classical black holes should be assigned temperature (and in general, thermodynamic features) (see Curiel, 2014).

κ and the observation of point coincidences exist (which run, in a non-redundant fashion, via the usage of a thermometer).

If case (2) was empirically established, even the empirical content linked to GR's formalism would be increased as correspondence rules between the surface gravity in GR and the sensation of heat are established for the first time. Note though that there seems to be an in principle limit to our human sensation of temperature—to still under-exaggerate—far off from the sensational sensitivity needed.

The problem, however, is that we expect from the QFT derivation of Hawking radiation that $T = \frac{\kappa}{8\pi}$ is too small for detection,[19] not to say directly perceivable by a human observer.[20]

What is the thermodynamic interpretation good for then? That the thermodynamic interpretation could in principle increase empirical access to GR is a necessary requirement for that black hole thermodynamics can be rendered an instance of thermodynamics proper. Therefore, that we can at all *conceive* of the analogy as potentially increasing empirical access to GR is a first good sign for that the analogy between black hole thermodynamics and thermodynamics proper is more than an analogy; in other words, our intuitive trust, if any, in black hole thermodynamics qua thermodynamics derives from the conceivability that thermometers, photon gases, ... can be brought into contact with a black hole.[21] But this means that it is also *considerations of increased empirical accessibility* of GR via thermodynamics which foster the idea that black hole thermodynamics is thermodynamics proper.

4 Guidelines From GR's Conceptual Interpretation

Practice in quantum gravity research—the search for a successor theory to GR—goes far beyond motivating successors to GR from exploring GR's empirical content. Different quantum gravity approaches are rather suggested from stressing specific conceptual takes on GR, that is: Issues of categorial interpretation—Is the theory a geometrised spacetime theory? Is the theory a hydrodynamic theory? Is the theory a causal theory?—provide different perspectives on the theory; these conceptual categorisations of GR then suggest natural starting points for extrapolating GR, say from how these conceptual categorisations are already known

[19]For astronomical black holes, the corresponding Hawking radiation is many orders smaller than the temperature of the cosmic microwave backgrounds. Small enough black holes for detection, on the other hand, would evaporate away too fast to be measurable.

[20]From the analogy at the level of GR, the putative black hole temperature can only be determined as equal to the surface gravity up to a constant pre-factor, which leaves the magnitude of the temperature undetermined.

[21]See also Prunkl and Timpson (2017).

to be linked to certain specific extrapolative strategies in other contexts.[22] For better illustration, let me provide some *examples* for links between conceptual interpretational stances on GR and extrapolative hypotheses:

- The categorisation of GR as a geometric spacetime theory directly derives from its standard differential geometric presentation (see, for instance, the canonical textbooks by Wald 2010 or Misner et al. 2017). It has been explicitly voiced as such, among others, by Friedman (2014) and Maudlin (2012).[23] In giving the gravitational interaction a special character as intrinsically geometric, the geometric viewpoint naturally motivates a semi-classical viewpoint according to which the geometric nature of spacetime is conserved and becomes the arena for quantum fields (semi-classical gravity).[24]
- GR can be seen as a locally thermodynamic theory: among other things,[25] the field equations can be interpreted as a balance equation of a heat flux through the horizon of a (local) Rindler observer (see Jacobson, 1995).[26] Taking the thermodynamic viewpoint seriously—including the stipulation of a generalised second law of thermodynamics[27]—entails the holographic principle, which itself suggests a hypothesis on the number of microstates within a volume as being bounded by their surrounding area (see, for instance, Bousso, 2002, section III).
- Connected to the previous viewpoint, GR can be seen as a hydrodynamic theory: under restriction to spacetimes with at least one Killing symmetry, the field equations take the form of the Navier–Stokes equations.[28] Rendering

[22]The two examples invoked for demonstrating how enlarging the concrete interpretation drives theory change can also be thought of as special cases of categorial interpretations, namely categorial interpretations which have (in-principle) empirical testability: In the case of demanding that the surface gravity can also be measured via thermometers, GR is conceptualised as a (partly) thermodynamic theory, and, in the case of enriching the chronometric access to GR, GR is conceptualised as a (locally) special relativistic theory in a sufficiently strong sense (not just the dynamical equations are locally Lorentz-invariant but they are form-invariant).

[23]The alternative position is that of the field view, which loosely speaking, renders the metric field as just one field, among others. Of course, the metric field has special properties but so does every other field; the basic methodological posit behind the field view is then not to mistake (arguably contingent) matters of representation for decisive facts. The particle physics approach to GR (“spin-2") as, for instance, promoted by Weinberg and Dicke (1973) (see Salimkhani, 2017 for a philosophical introduction), and hydrodynamic viewpoints on GR (see below) promote exactly such kind of take. Within the philosophy of physics, the field view is first and foremost promoted by adherents to the dynamical approach (see Brown, 2005 as the locus classicus).

[24]For a discussion of arguments against semi-classical accounts as fundamental theories (see Callender and Huggett, 2001 and Wüthrich (2005)). For a modern thought experiment for testing the scope of a semi-classical gravity paradigm (see Bose et al., 2017).

[25]See Padmanabhan (2016), section 1 for a comparison of various account of gravity as a thermodynamic theory.

[26]Strictly speaking, this requires, however, adherence to the Unruh effect, that is an effect from QFT in flat spacetime.

[27]The second law holds for matter and black hole entropy in total.

[28]See, for instance, Rodrigues Jr. and de Oliveira (2016, chapter 15).

the field equations as hydrodynamic equations suggests considering fluctuation corrections to the field equations. This is exactly what is done in Hu's stochastic gravity (see Hu 1999).

- Following a theorem by Malament (1977), spacetime structure in GR can be split up into (continuous) causal structure and local volume information. By making the causal viewpoint central, the volume information can be demoted to a secondary feature. (Provided that we assume a finite number of causal events, the volume information can then be obtained through counting the number of causal events in a region.) This naturally leads to causal set theory.
- GR is universal coupling. This can be seen either as a necessary feature of its putative geometric nature (see first point), or as a sign for that it is the result of coarse-graining (its geometric nature is then rather a representational coincidence).[29] The latter view suggests taking analogies between GR and non-geometrical theories more seriously again.

We can note that, in practice, many principles and associated viewpoints just seem to make *themselves* remarkable, say by analogy to other theories. However, one way to systematically work out decisive principles and thus viewpoints of GR is a contrastive approach where GR is compared to neighbouring spacetime theories in order to reveal features making GR special, and thus worth specific attention—see Lehmkuhl et al. (2017) for a project along these lines.

Furthermore, we can note that each conceptual take on GR above can usually be backed up from several strands of reasoning. It is an urgent question then for the philosophy of quantum gravity qua philosophy of discovery to what extent robustness arguments which work at an entirely conceptual and thus non-empirical level can support the pursuit-worthiness or even plausibility of a specific conceptual viewpoint on GR.

Concerning the motivation of hypotheses from a specific viewpoint, we can make the following two observations: (1) Each of these viewpoints on GR suggests characteristic extrapolative hypotheses. In many cases, these hypotheses are suggested by analogy: a particular viewpoint is known to be linked to specific extrapolative strategies in the context of other theories, which can be exploited in the context of GR then. In some other cases, taking a viewpoint seriously more or less requires making new technical commitments or discarding old ones. Causal set theory is a good example for this as it arises from taking the causal viewpoint seriously which means discretising the space of events, and thus giving up a notion of Lorentz-symmetry and Lorentzian manifold at high energies.[30] (2) By suggesting characteristic extrapolative hypotheses, categorial interpretations obtain a decisive role in motivating prima facie independent approaches. What we call the principles of a specific approach to quantum gravity is thus often already rooted in a specific way of looking at GR, i.e. a specific categorial interpretational stance.

---

[29] See Feynman et al. (2003, section 1.5.)

[30] Which does not mean that CST is inconsistent with Lorentz symmetry at lower energies (see Dowker et al., 2004).

5 Conclusion

This chapter showed how both empirical and conceptual interpretations of GR give straightforward suggestions for a successor theory to GR. I identified the following heuristic rules for driving theory change from interpreting GR:

- At the level of the empirical interpretation:
  1. Take the empirical content seriously at a general level: Look for entities adhered to in the theory but knowingly described more accurately in certain domains by other theories. This suggests potential directions for unification.
  2. Take the empirical content seriously at a specific level: Look for extratheoretical measurement methods of otherwise empirically uninterpreted parts of the theory. This suggests potential directions for unification of GR.
- At the level of the conceptual interpretation:
  1. Explore possible conceptual readings of the theory. Directions for conceptual readings are at first suggested by striking conceptual analogies to other theories, and prominent conceptual features or results as such.
  2. Try extrapolative schemes generally associated to a certain conceptual categorisation of the theory.
  3. If a conceptual viewpoint is novel, that is not known from other theories (as the viewpoint of GR as a causal theory), explore options for making this viewpoint centre-stage nevertheless. This may easily involve violating principles of the current theory (such as that of a Lorentzian manifold in the case of causal set theory). In fact, possible conflicts again drive progress here, as they provide a concrete problem to tackle.

Now, the heuristic of using empirical and concrete interpretation as a guide towards successor theories can, of course, again be used on the thereby suggested successor theories such as QFT in curved spacetime, semi-classical gravity, and quantum GR. The full merit of the promoted methodology thus derives from its iterative applicability.

A final remark is in order on the relationship between theory interpretation and internal problems as theory drivers: The chapter investigated how theory interpretation suggests directions for theory change. At the same time, it is a common theme—in particular, in quantum gravity research where the problem is first of all theoretical and not empirical[31]—that internal problems suggest the alteration of a theory into a certain direction. Now, it is worth stressing that of course not the problem as such but rather the conceptual interpretation of the theory

[31]Examples include: (1) GR spacetimes are partly singular. These spacetimes are not sufficiently predictive. (2) QFT in curved spacetime lacks back-reaction (which we know is desirable from GR). (3) Semi-classical gravity is possibly incoherent and thus at least predictively limited. (4) Perturbative quantisation of gravity leads to an only effectively renormalisable, that is predictively limited theory.

featuring the problem suggests the direction for theory change. The problem as such can at most only indicate the need for theory change.

Acknowledgments I would like to thank Christian Wüthrich, James Read, Kian Salimkhani, Manus Visser, Patrick Dürr, and Rasmus Jaksland for earlier comments on this manuscript. I would in particular like to thank James Read for discussions on EPS and radar coordinates, which are picked up on in detail in a joint manuscript in progress. Note that this paper is an extract from my PhD thesis "Philosophy of quantum gravity as a philosophy of discovery" which was funded by the Swiss National Science Foundation (SNF) under the project number 105212_165702. The content of this work is solely the responsibility of the author and does not represent the official views of the SNF. I am very thankful towards the SNF for their generous support.

References

John L. Bell and Herbert Korté. Hermann Weyl. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

Sougato Bose, Anupam Mazumdar, Gavin W. Morley, Hendrik Ulbricht, Marko Toroš, Mauro Paternostro, Andrew A. Geraci, Peter F. Barker, M.S. Kim, and Gerard Milburn. Spin entanglement witness for quantum gravity. *Phys. Rev. Lett.*, 119(24):240401, 2017.

Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825, 2002.

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

Otávio Bueno. Empirical adequacy: A partial structures approach. *Studies in History and Philosophy of Science Part A*, 28(4):585–610, 1997.

Craig Callender and Nick Huggett. *Physics meets philosophy at the Planck scale: Contemporary theories in quantum gravity*. Cambridge University Press, 2001.

Rudolf Carnap. *Philosophical foundations of physics*, volume 966. Basic Books New York, 1966.

Robert Alan Coleman and Herbert Korte. Jet bundles and path structures. *Journal of Mathematical Physics*, 21(6):1340–1351, 1980.

Erik Curiel. General relativity needs no interpretation. *Philosophy of Science*, 76(1):44–72, 2009.

Erik Curiel. Classical black holes are hot. *arXiv preprint arXiv:1408.3691*, 2014.

Sebastian de Haro and Henk W. de Regt. Interpreting theories without a spacetime. *European Journal for Philosophy of Science*, pages 1–40, 2018.

John Dougherty and Craig Callender. Black hole thermodynamics: More than an analogy? In A. Iljas & B. Loewer, editors, Guide to the *Philosophy of Cosmology*. Oxford University Press (forthcoming). http://philsci-archive.pitt.edu/13195/.

Fay Dowker, Joe Henson, and Rafael D. Sorkin. Quantum gravity phenomenology, lorentz invariance and discreteness. *Modern Physics Letters A*, 19(24):1829–1840, 2004.

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

Richard P. Feynman, Fernando B. Morinigo, and William G. Wagner. Feynman lectures on gravitation, 2003.

Samuel C. Fletcher. Light clocks and the clock hypothesis. *Foundations of Physics*, 43(11):1369–1383, Nov 2013. ISSN 1572-9516. URL https://doi.org/10.1007/s10701-013-9751-3.

Michael Friedman. The scientific image by Bas C. van Fraassen. *The Journal of Philosophy*, 79 (5):274–283, 1982.

Michael Friedman. *Foundations of space-time theories: Relativistic physics and philosophy of science*, volume 113. Princeton University Press, 2014.

Hans Halvorson. The semantic view, if plausible, is syntactic. *Philosophy of Science*, 80(3):475–478, 2013.

Stephen W. Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

B. L. Hu. Stochastic gravity. *International Journal of Theoretical Physics*, 38(11):2987–3037, 1999.

Ted Jacobson. Thermodynamics of spacetime: the Einstein equation of state. *Phys. Rev. Lett.*, 75 (7):1260, 1995.

Claus Kiefer. Quantum gravity. *Int. Ser. Monogr. Phys.*, 124:1–308, 2004.

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.

Dennis Lehmkuhl. Mass–energy–momentum: Only there because of spacetime? *The British Journal for the Philosophy of Science*, 62(3):453–488, 2011.

Dennis Lehmkuhl. Literal versus careful interpretations of scientific theories: The vacuum approach to the problem of motion in general relativity. *Philosophy of Science*, 84(5):1202–1214, 2017.

Dennis Lehmkuhl, Gregor Schiemann, and Erhard Scholz. *Towards a theory of spacetime theories*, volume 13. Springer, 2017.

Niels Linnemann. On interpreting the empirical content of general relativity. *Unpublished draft*, 2019.

Sebastian Lutz. What's right with a syntactic approach to theories and models? *Erkenntnis*, 79(8): 1475–1492, 2014.

Sebastian Lutz. What was the syntax-semantics debate in the philosophy of science about? *Philosophy and Phenomenological Research*, 95(2):319–352, 2017.

David B. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of mathematical physics*, 18(7):1399–1404, 1977.

David B. Malament. *Topics in the foundations of general relativity and Newtonian gravitation theory*. The University of Chicago Press, 2012.

Tim Maudlin. *Philosophy of physics: Space and time*, volume 5. Princeton University Press, 2012.

Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. Princeton University Press, 2017.

Thanu Padmanabhan. Exploring the nature of gravity. *arXiv preprint arXiv:1602.01474*, 2016.

Volker Perlick. On the radar method in general-relativistic spacetimes. In *Lasers, Clocks and Drag-Free Control*, pages 131–152. Springer, 2008.

Carina Prunkl and Christopher Timpson. Black Hole Entropy is Entropy and not (necessarily) Information. *Unpublished draft*, 2017.

Guillaume Rochefort-Maranda. Constructive empiricism and the closure problem. *Erkenntnis*, 75 (1):61–65, 2011.

Waldyr A. Rodrigues Jr. and Edmundo Capelas de Oliveira. *The Many Faces of Maxwell, Dirac and Einstein Equations: A Clifford Bundle Approach*, volume 922. Springer, 2016.

Kian Salimkhani. Quantum Gravity: A Dogma of Unification? *Philosophy of Science – Between Natural Science, the Social Sciences, and the Humanities, European Studies in Philosophy of Science, Springer*, 2017.

Howard Stein. Some reflections on the structure of our knowledge in physics. In *Logic, Metholodogy and Philosophy of Science, Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science*, pages 633–55. Citeseer, 1994.

Frederick Suppe. Understanding scientific theories: An assessment of developments, 1969–1998. *Philosophy of Science*, 67:S102–S115, 2000.

John Lighton Synge. *Relativity: the general theory*. North-Holland Publishing Company Amsterdam, 1960.

Bas C. Van Fraassen. *The scientific image*. Oxford University Press, 1980.

Robert M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics*. The University of Chicago Press, 1994.

Robert M. Wald. *General relativity*. The University of Chicago Press, 2010.

David Wallace. The case for black hole thermodynamics, Part i: phenomenological thermodynamics. *arXiv preprint arXiv:710.02724*, 2017.

James Owen Weatherall. Against dogma: On superluminal propagation in classical electromagnetism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 48:109–123, 2014. ISSN 1355-2198. doi:https://doi.org/10.1016/j.shpsb.2014.08.005. URL http://www.sciencedirect.com/science/article/pii/S1355219814000896. Relativistic Causality.

Steven Weinberg and R. H. Dicke. *Gravitation and cosmology: principles and applications of the general theory of relativity*. Wiley New York, 1973.

Hermann Weyl. Zur infinitesimalgeometrie: Einordnung der projektiven und der konformen auffasung. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1921:99–112, 1921.

Rasmus Grønfeldt Winther. The structure of scientific theories. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

Christian Wüthrich. To quantize or not to quantize: fact and folklore in quantum gravity. *Philosophy of Science*, 72(5):777–788, 2005.

Christian Wüthrich. Are black holes about information?, 2019. In Richard Dawid, Radin Dardashti, and Karim Thébault, editors, *Epistemology of Fundamental Physics*, Cambridge University Press, 2019.

Explanation, Geometry, and Conspiracy in Relativity Theory

James Read

Abstract I discuss the debate between *dynamical* versus *geometrical* approaches to spacetime theories, in the context of both special and general relativity, arguing that (a) the debate takes a substantially different form in the two cases; (b) different versions of the geometrical approach—only some of which are viable—should be distinguished; (c) in general relativity, there is no difference between the most viable version of the geometrical approach and the dynamical approach. In addition, I demonstrate that what have previously been dubbed two 'miracles' of general relativity admit of no resolution from within general relativity, on either the dynamical or 'qualified' geometrical approaches, modulo some possible hints that the second 'miracle' may be resolved by appeal to recent results regarding the 'geodesic principle' in GR.

1 Introduction

It is roughly a decade since the groundbreaking work of Brown (2005), Brown and Pooley (2001, 2006) brought into the mainstream philosophy of physics literature the debate between *dynamical* versus *geometrical* approaches to spacetime theories. At the most general level, this debate centres upon the following question: *whence the chronogeometric significance of the metric field?* That is, why is the metric field (in theories such as special and general relativity) surveyed by rods and clocks built from matter fields? While advocates of the geometrical approach maintain that the metric field (in some sense) *explains* or *constrains* the form of the dynamical laws for matter fields, such that those fields behave such as to survey the metric field, advocates of the dynamical approach, by contrast, claim that an account of the chronogeometric significance of the metric field may begin from considerations regarding only the dynamical laws governing matter fields themselves.

J. Read (✉)
Pembroke College, Oxford, UK
e-mail: james.read@philosophy.ox.ac.uk

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_9

Of course, this is vague; in Sect. 3 of this paper, I sharpen significantly the above presentation of the debate. Nevertheless, even at this early stage, a number of genuine and substantial questions arise. Some of those which concern me in this paper are the following:

1. Does the dynamical/geometrical debate take the same form in the context of theories with fixed metric structure (such as special relativity (SR)) as it does in theories with dynamical metric structure (such as general relativity (GR))?
2. What notion of explanation is at play in this debate? Does answering this question reveal multiple different senses in which the dynamical/geometrical approaches may be understood?
3. Are the dynamical and geometrical approaches truly distinct from one another at all?
4. How does the dynamical approach relate to e.g. the 'spacetime functionalism' of Knox (2017, 2013, 2014), or recent discussion on these matters by Weatherall? (Weatherall, 2017, §6)

In brief, my answers run as follows. On (1), I argue in Sect. 3 of this paper that there exist *significant* differences regarding this debate as it occurs in the context of SR, versus as it occurs in GR. The principal reason for this is that, while the advocate of the dynamical approach may be regarded as seeking to *ontologically reduce* the metric field in theories with fixed metric structure (such as SR) to the symmetry properties of matter fields (cf. Brown and Read 2020; Myrvold 2017), she does *not* attempt to make such a move in theories with dynamical metric structure, such as GR.[1]

On (2), I argue in Sect. 4 that it is important to distinguish between what I call 'qualified' versus 'unqualified' explanations in the context of this debate.[2] Once this distinction is made, the geometrical approach bifurcates into two positions, which I call, respectively, the 'qualified' and 'unqualified' geometrical approaches. In Sect. 5, I argue that distinguishing between these two positions is crucial, for while the former version of the geometrical approach is tenable, the latter is *not*;

[1]For further discussion regarding how this debate changes on moving from SR to GR, see Brown (2005), Brown and Read (2020), Read et al. (2018). In light of the fact that the advocate of the dynamical approach does not attempt to undertake an ontological reduction of the above-described kind in the context of GR, one might be inclined to conclude: 'So much the worse for the dynamical approach in the context of GR, as a distinct view in the landscape'. Below, I will argue that there is something to this concern, for (I maintain) there is no difference in the GR context between the dynamical approach and the most defensible version of the geometrical approach.

[2]It is worth flagging that I will offer these two notions of explanation without claiming (or seeking) to give a full conceptual analysis of the notion of scientific explanation; in my view, the distinction between 'qualified' and 'unqualified' explanations is still a valuable and comprehensible one (providing, as I see it, at least some of the 'explanatory concepts' which Norton suggests may be necessary for 'a full understanding of constructivism [i.e., the dynamical approach]' (Norton 2008, p. 824)), even in the absence of such an analysis. (In this regard, cf. the methodology of Weatherall 2017, pp. 15–16.)

nevertheless, these two views have been run together in much of the literature on this topic up to this point.

On (3), I argue in Sect. 5 that, while the 'unqualified' version of the geometrical approach is distinct from the dynamical approach in the context both of theories such as SR and of theories such as GR, the 'qualified' geometrical approach, by contrast, is only distinct from the dynamical approach in the former context.

On (4), I argue that there is an important sense in which Knox's spacetime functionalism, according to which 'the spacetime role is played by whatever defines a structure of local inertial frames' (Knox, 2017, p. 22), constitutes an *extension* of the dynamical approach—in essence stating that whichever structure has chronogeometric significance may be identified as playing the functional role of spacetime, and therefore, on a functionalist approach to the definition of physical quantities, may be identified as being spatiotemporal *tout court*.[3] In addition, I argue that Weatherall (2017, §6) is most plausibly read as both (a) embracing spacetime functionalism, and (b) embracing either the dynamical or the 'qualified' geometrical approach.

Along the way, a number of other tasks are accomplished. Most notably, I demonstrate that what were labelled in Read et al. (2018, §5) two 'miracles' of GR—(1) that all dynamical laws for matter fields have the same local (Poincaré) symmetry properties; and (2) that these local symmetries coincide (in the relevant regime in which curvature effects may be ignored) with the symmetries of the ontologically autonomous metric field in the theory—admit of no resolution from within GR, on any plausible form of the dynamical or geometrical approaches, modulo some hints from recent work on the so-called geodesic principle in GR regarding the second 'miracle'.[4]

2 Background

Before proceeding to the matters outlined above, I review in this section some standard discussion regarding (1) the formulation of classical spacetime theories (Sect. 2.1); (2) symmetries in such theories (Sect. 2.2); and (3) presentations of special and general relativity (Sect. 2.3).

[3]By contrast, there is a sense in which advocates of the dynamical approach need not speak of 'spacetime' at all—cf. Brown and Read (2020, §3.1).

[4]It is worth noting that these two 'miracles' of GR may admit of resolution in a *successor* theory to GR, in a manner analogous to that in which the 'miracle' of the coincidence of gravitational and inertial masses in Newtonian mechanics was resolved on moving to GR. See Weatherall (2011a) for a detailed discussion of the explanation of the coincidence of gravitational and inertial masses in GR, and Read (2019) for how the two 'miracles' of GR may be resolved on moving to one particular successor theory—viz., perturbative string theory.

2.1 Spacetime Theories

Let us say—following e.g. Anderson (1967), Pooley (2013), Pooley (2017), Thorne et al. (1973)—that the *kinematically possible models* (KPMs) of a given spacetime theory are picked out by tuples $\langle M, \Phi_1, \ldots, \Phi_n \rangle$, with (a) M a (four-dimensional) differentiable manifold; and (b) the $\Phi_1, \ldots, \Phi_n$ various (tensor) fields on M.[5] Given a class of KPMs for a given theory, let us then say that the *dynamically possible models* (DPMs) of that theory are those KPMs the $\Phi_1, \ldots, \Phi_n$ of which satisfy certain specified dynamical equations.

To illustrate, consider two examples. First, take a special relativistic massless Klein–Gordon theory (call it **KGS**). In this theory, KPMs are triples $\langle M, \eta_{ab}, \varphi \rangle$, where η_{ab} is a fixed Minkowski metric field on M (fixed *identically* in all KPMs—see Pooley 2017, p. 115), and φ is a real scalar field on M. DPMs of **KGS** are picked out as those KPMs the fields of which satisfy the massless Klein–Gordon equation,[6]

$$\eta_{ab} \nabla^a \nabla^b \varphi = 0. \tag{1}$$

As a second example, consider a *general* relativistic Klein–Gordon theory (call it **KGG**). In this case, KPMs are again triples $\langle M, g_{ab}, \varphi \rangle$—this time, however, g_{ab} is a generic Lorentzian metric field on M, *not* fixed in all DPMs of the theory. In this case, DPMs are picked out by the GR Klein–Gordon equation,[7]

$$g_{ab} \nabla^a \nabla^b \varphi = 0, \tag{2}$$

and the Einstein field equations,[8]

$$G_{ab} = 8\pi T_{ab}, \tag{3}$$

where T_{ab} is the stress-energy tensor associated with φ.[9]

[5]In principle, we should not exclude other types of field on M—e.g. spinor fields; pseudotensors; tensor densities; etc. (For arguments for taking these latter two classes of object seriously, see Pitts 2006, 2010.) In this paper, however, I focus exclusively upon the case in which the Φ_i are tensor fields.

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

[7]The torsion-free derivative operator ∇_a now compatible with g_{ab}, so that $\nabla_a g_{bc} = 0$.

[8]These are the Einstein field equations with vanishing cosmological constant Λ. For $\Lambda \neq 0$, the field equations read $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$.

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

2.2 Symmetries

I now draw a standard distinction between *metric symmetries*, and *dynamical symmetries* (cf. e.g. Earman 1989, §3.4).

For a given metric field, let us say that a coordinate transformation is a *metric symmetry* (sometimes: an *isometry*) just in case the metric field is unaltered by the coordinate transformation. For example, the symmetries of the Minkowski metric field η_{ab} of special relativity are the *Poincaré transformations*—those affine transformations[10]

$$x^{\mu} \rightarrow \Lambda^{\mu}{}_{\mu'} x^{\mu'} + a^{\mu} \tag{4}$$

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

$$\Lambda_{\mu'}{}^{\mu} \Lambda_{\nu'}{}^{\nu} \eta_{\mu\nu} = \eta_{\mu'\nu'}, \tag{5}$$

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

In addition to the notion of metric symmetries, it is useful to introduce the notion of a *dynamical symmetry*. A coordinate transformation is a *dynamical symmetry* just in case the dynamical equations governing non-gravitational fields take the same form in coordinate systems related by that transformation.[11] For example, transforming the SR Klein–Gordon equation (1) from one coordinate system to a second via an affine transformation (where in this case $M^{\mu}{}_{\mu'}$ represents the linear transformation matrix)[12]

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

[11] In this paper, I mean by 'matter fields', or 'non-gravitational fields', those for which there exists an associated stress-energy tensor, and by 'gravitational fields' those for which there exists no such stress-energy tensor—this distinction is in the spirit of Lehmkuhl (2011). In the context of GR, this means that the metric field is identified as a gravitational field, whereas all other fields typically of interest (e.g., Klein–Gordon fields, electromagnetic fields, etc.) are matter fields. Clearly, there exist subtle issues here regarding the possibility of defining a stress-energy tensor associated with the metric field in GR—see Curiel (2018) for a proof against this possibility, and e.g. Hoefer (2000), Lam (2011), Read (2018) for related discussion. Note also that this distinction between matter and gravitational fields may break down in the case of other spacetime theories—for example, in Newtonian gravitation theory (cf. Sect. 5.2.2), it is possible to define a stress-energy tensor associated with the potential φ, in spite of this field naturally being regarded as 'gravitational' (cf. Dewar and Weatherall 2018). Nevertheless, for my purposes, the above distinction will suffice.

[12] Here, again, I use a coordinate-based description. Note that I do not transform the fixed fields—cf. Pooley (2017, p. 115).

$$\eta_{\mu\nu}\nabla^{\mu}\nabla^{\nu}\varphi = 0 \tag{6}$$

$$\longrightarrow \quad \eta_{\mu\nu}M^{\mu}{}_{\mu'}M^{\nu}{}_{\nu'}\nabla^{\mu'}\nabla^{\nu'}\varphi = 0, \tag{7}$$

one finds that such an equation is invariant under the transformation when (5) is satisfied—i.e. if the affine transformation is a Poincaré transformation. Thus, the dynamical symmetries associated with (1) at least include the Poincaré transformations.

2.3 *Special and General Relativity*

Having introduced the necessary details regarding spacetime theories and their symmetries, in this section I characterise—with both greater precision and generality—what it means for a given theory to be 'special relativistic' (Sect. 2.3.1), versus 'general relativistic' (Sect. 2.3.2).

2.3.1 Special Relativity

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

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

The latter criterion is referred to by Brown as the *big principle*—see e.g. Brown (2005, §8.4.1).[13] Note that, by construction, metric and dynamical symmetries coincide in special relativistic theories.

2.3.2 General Relativity

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

- KPMs at least include a Lorentzian metric field g_{ab}—so may be written $\langle M, g_{ab}, \Phi_1, \ldots, \Phi_n \rangle$.

[13]In his 1908 paper Minkowski (1909), Minkowski referred to this principle as the *world-postulate*—for discussion, see Brown (2005, §8.1).

- DPMs are picked out by dynamical equations for the $\Phi_1, \ldots, \Phi_n$, along with the Einstein field equations $G_{ab} = 8\pi T_{ab}$, where T_{ab} is the stress-energy tensor for the $\Phi_1, \ldots, \Phi_n$.

This characterisation of a general relativistic theory is very weak—note, in particular, that there is no guarantee in general relativistic theories so understood that metric symmetries coincide locally with dynamical symmetries (as was the case for special relativistic theories, as presented above).[14] Such requirements may be imposed via restriction to those models of GR which satisfy further conditions; from the point of view of the matter of symmetry coincidence, one particular auxiliary condition which will be of significance is:[15]

- Instantiation of the *strong equivalence principle* (SEP).

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

The SEP is intended to capture facts regarding the 'local validity' of SR in GR. Brown puts the point thus:[16]

> There exists in a neighbourhood of each event preferred coordinates, called *locally inertial* at that event. For each fundamental non-gravitational interaction, to the extent that tidal gravitational forces can be ignored, the laws governing the interaction find their simplest form in these coordinates. This is their *special relativistic form*, independent of spacetime location (Brown, 2005, p. 169).

Here, there exist a number of subtleties regarding what is meant by the qualification 'to the extent that tidal gravitational forces can be ignored', and moreover regarding whether other foundational principles GR—for example *minimal coupling*, which is a heuristic prescription for the construction of dynamical laws for non-gravitational fields in GR from those in SR—are compatible with the SEP as formulated above. Since these matters are not directly relevant to my purposes in this paper, I refer the reader to Brown and Read (2016, 2020), Read et al. (2018) for detailed discussion. For today, the essential aspect of the SEP is the imposition that, in the neighbourhood of any $p \in M$ in GR, laws of physics recover their 'special relativistic form'—where I shall understand this to mean: *a Poincaré invariant form*. Clearly, this is a particular restriction on the matter sector in the theory.

---

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

[15]Other conditions which one may be interested in imposing upon the class of GR solutions in which one is interested are e.g. *energy conditions*, for such conditions are often understood to be tied to the restriction to 'physically reasonable' matter (for example, to conditions that energy cannot be negative). For a recent virtuoso study of energy conditions, see Curiel (2017).

[16]Other similar presentations of the SEP can be found in e.g. Knox (2013, §3.4) and Knox (2014, p. 874).

To illustrate, consider again the general relativistic Klein–Gordon equation (2). Written in an arbitrary coordinate basis, this reads

$$g_{\mu\nu}\partial^{\mu}\partial^{\nu}\varphi + \Gamma^{\mu}{}_{\nu\mu}\partial^{\nu}\varphi = 0. \tag{8}$$

Recall that in a coordinate basis $\{e_\mu\}$, the *connection components* $\Gamma^{\mu}{}_{\nu\rho}$ associated with a derivative operator ∇_a are defined by $\nabla_\rho e_\nu =: \Gamma^{\mu}{}_{\nu\rho} e_\mu$. Then, at any $p \in M$, we can choose *normal coordinates*, such that $\Gamma^{\mu}{}_{(\nu\rho)}(p) = 0$ in those coordinates; for a torsion-free derivative operator, we can in fact choose normal coordinates such that $\Gamma^{\mu}{}_{\nu\rho}(p) = 0$. (Note that the connection components away from p will in general *not* vanish.) If the unique torsion-free, metric compatible derivative operator is used, then in normal coordinates we also have $\partial_\rho g_{\mu\nu}(p) = 0$, and one may further restrict to the subset of normal coordinates at p such that $g_{\mu\nu}(p) = \text{diag}(-1, 1, 1, 1)$. Since $g_{\mu\nu}(p)$ takes this diagonal form—preserved under Poincaré transformations—in this restricted class of normal coordinates at p, one might write $g_{\mu\nu}(p) = \eta_{\mu\nu}$ (cf. e.g. Misner et al. 1973, p. 1055). This notwithstanding, however, any claim to the effect that the metric field 'reduces' to the Minkowski metric at p in normal coordinates should be met with suspicion—for in general, second (and higher) order derivatives of the metric field *do not* vanish at p, in these coordinates.

In normal coordinates at $p \in M$, (2) (in a coordinate basis, (8)) takes a particularly simple form at p:

$$\eta_{\mu\nu}\partial^{\mu}\partial^{\nu}\varphi = 0; \tag{9}$$

moreover, this form (with the metric diagonalised) is retained in all frames related by Poincaré transformations. This illustrates the sense in which certain dynamical equations for non-gravitational fields recover locally a Poincaré invariant form. That *all* dynamical laws for non-gravitational fields in GR manifest this quality is a statement of the SEP. Importantly, note that, absent the imposition of the SEP, it is *not* the case that all dynamical equations for matter fields in GR need be locally Poincaré invariant. For example, there exists no *a priori* prohibition on the existence of matter fields obeying dynamical laws which are locally *Galilean* invariant, in a spacetime theory with a dynamical, Lorentzian metric field satisfying the Einstein field equations.[17]

Why should one restrict to those solutions of GR in which the SEP is satisfied? The reason is that this principle—which ensures that, locally, the (Poincaré) symmetries of the metric field[18] coincide with the (Poincaré) symmetries of the dynamical laws governing matter fields—is typically regarded to constitute an

[17] See Knox (2013), Read (2019) for discussion of this possibility.

[18] Again, modulo subtle issues regarding the qualification 'to the extent that tidal gravitational forces can be ignored'—see Brown and Read (2016), Brown and Read (2020), Read et al. (2018) for discussion.

important condition for the *chronogeometricity* of the metric field—that is, for intervals as given by the metric field to be read off by stable rods and clocks built from matter fields. As Brown writes,[19]

> It is because of . . . local Lorentz covariance that rods and clocks, built out of the matter fields which display that symmetry, behave as if they were reading aspects of the metric field and in so doing confer on this field a geometric meaning. That light rays trace out null geodesics of the field is again a consequence of the strong equivalence principle, which asserts that locally Maxwell's equations of electrodynamics are valid (Brown, 2005, p. 176).

That is, the SEP is, it is argued, an important condition for the metric field g_{ab} in GR to have (local) *operational meaning*.[20,21] I return to discuss further the SEP, in light of the so-called *geodesic principle* in GR, in Sect. 6.

3 The Dynamical/Geometrical Debate

The above in hand, in this section I demonstrate how the dynamical/geometrical debate plays out in the context of both SR (Sect. 3.1) and GR (Sect. 3.2);[22] a detailed reconsideration of the geometrical approach will follow in Sects. 4 and 5. At the most general level, the dynamical/geometrical debate centres upon the following question:

> Whence the metric field's chronogeometric significance?

Taking, as elaborated above, the (local) coincidence of metric and dynamical symmetries to be an important condition which must be fulfilled in one important means via which the metric field acquires its chronogeometric significance (viz., via

---

[19] One might reasonably pause over whether the 'as if' in the following passage is necessary, on Brown's account.

[20] One might wonder whether satisfaction of the SEP should be regarded as being a *necessary* condition for the metric field to have its chronogeometric significance, or rather as being a *sufficient* condition, or rather something else. In Read et al. (2018, pp. 15–16), it is indeed claimed that the SEP constitutes a necessary condition for chronogeometricity; however, it is perhaps more conservative to state that, with auxiliary assumptions such as the existence of stable rods and clocks, it constitutes a *jointly* sufficient condition for chronogeometricity. In this way, one does not rule out other possible means of gaining operational access to the metric field—for example, by using test particles which traverse null and timelike geodesics to gain access to conformal and projective structure, from which (via Weyl's theorem-type reasoning—cf. Ehlers et al. (1972), Weyl (1921), Weyl (1923), and with certain further additional assumptions) one can recover metric structure. There remains much further work to be done in order to understand fully these alternative means of gaining operational access to the metric field; cf. footnote 55 for some further discussion of Weyl's theorem, and Butterfield (2007, §4) for related discussion.

[21] Similarly, one might argue that postulating that metric symmetries coincide with dynamical symmetries in SR is an important condition for the metric field η_{ab} to have operational meaning in that case.

[22] There is a sense in which the lessons of Sects. 3.1 and 3.2 can be generalised to *all* theories with, respectively, fixed versus dynamical metric structure—see Brown and Read (2020, §5).

the SEP), one antecedent question which one might seek to address in order to answer the above is the following:[23]

> Why do metric symmetries coincide (locally) with dynamical symmetries?[24]

It is upon this latter question that much of the dynamical/geometrical debate has focussed. *Prima facie*, advocates of the *dynamical approach* (developed in particular by Brown 2005; Brown and Pooley 2001, 2006) appear to offer very different accounts of this coincidence of symmetries to advocates of the *geometrical approach* (for example, Friedman 1983 or Maudlin 2012). In the remainder of this section, I discuss the dynamical/geometrical debate in the context of both SR (Sect. 3.1) and GR (Sect. 3.2).

3.1 Special Relativity

In this subsection, I consider the account of the coincidence of metric and dynamical symmetries in SR proffered on the part of advocates of the geometrical (Sect. 3.1.1) and dynamical (Sect. 3.1.2) approaches.

3.1.1 The Geometrical Approach

Why, in special relativistic theories, do dynamical symmetries coincide with symmetries of the Minkowski metric field η_{ab}? Advocates of the *geometrical approach* to spacetime theories seek to answer this question via some appeal to η_{ab} itself. To be specific, in this paper I focus upon a version (later: versions) of the approach according to which the Minkowski metric field η_{ab} of SR is *ontologically autonomous and primitive*, and (somehow; in some sense to be cashed out) *constrains* the possible form of dynamical equations for matter, such that metric symmetries coincide with dynamical symmetries. As Maudlin (2012, pp. 117–8) writes,

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

[23]For the time being, my focus is on this mode of gaining operational access to the metric field—though I concede that there may be other means, as discussed in footnote 20 above, and in Sect. 6 below.

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

That a notion of constraint is at play on this view is manifest in the italicised portion of the above quotation. While advocates of the dynamical approach often object that such a notion of constraint or explanation is mysterious—for example, Brown writes 'It is wholly unclear how this geometrical explanation is supposed to work.' Brown (2005, p. 134)—I will assess in Sect. 4 the extent to which such objections find their mark. In the meantime, I turn to the dynamical approach to SR.

3.1.2 The Dynamical Approach

The dynamical approach offers a very different perspective on the coincidence of metric and dynamical symmetries in SR. According to this view, the metric field η_{ab} is not ontologically autonomous and primitive; rather, it is a *codification* of the symmetry properties of the dynamical equations governing matter fields. (One may, therefore, understand the dynamical approach to SR—and to theories with fixed metric structure more generally—as an ontological thesis; as a form of *relationalism*—cf. Pooley 2013, §6.3.2.) As Brown puts it: (Cf. also Brown and Pooley 2006, p. 80.)

> The appropriate structure is Minkowski geometry *precisely because* the laws of physics of the non-gravitational interactions are Lorentz covariant (Brown, 2005, p. 133).

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

> [T]he connection between spacetime [metric] structure and dynamical symmetries and asymmetries is analytic (Myrvold, 2017, p. 13).

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

The question of whether the dynamical approach to SR is viable has been widely discussed—see e.g. Acuña (2016), Brown (2005), Huggett (2009), Janssen (2009), Norton (2008), Pooley (2013), Stevens (2015, 2017). In this paper, I focus on a different issue: whether advocates of the dynamical approach have been fair to the geometrical approach, and whether 'geometricians' can, in fact, offer a coherent answer to the question of why metric and dynamical symmetries coincide, in SR. Before doing so, however, I consider how the nature of the dynamical/geometrical debate shifts on moving to GR.

3.2 General Relativity

In the GR context, advocates of both the dynamical and geometrical approaches *agree* that the metric field g_{ab} is an ontologically autonomous entity, obeying its own dynamical equations, and not straightforwardly reducible to (symmetries of dynamical equations governing) matter fields, as per the dynamical approach to SR.[25] However, the two approaches *prima facie* continue to issue different verdicts on the question of why metric and dynamical symmetries may be taken (locally) to coincide. In this subsection, I review the geometrical (Sect. 3.2.1) and dynamical (Sect. 3.2.2) approaches to GR.

3.2.1 The Geometrical Approach

Advocates of the version (later: versions) of the geometrical approach to GR under consideration in this paper maintain that, locally in the neighbourhood of any $p \in M$ (and in the regime in which 'tidal gravitational forces' may be ignored—cf. Sect. 2.3.2), metric and dynamical symmetries coincide (in accordance with the SEP), because the metric field g_{ab} (somehow; in some sense to be cashed out) *constrains* the possible form of dynamical equations for matter, such that metric symmetries coincide (locally) with dynamical symmetries. While, again, the advocate of the dynamical approach may find the notion of constraint here mysterious, I discuss in Sects. 4 and 5 the extent to which these matters can be accounted for by advocates of the geometrical approach.

3.2.2 The Dynamical Approach

Assuming that the metric field in GR is not ontologically reducible to (symmetries of dynamical laws governing) matter fields, the foregoing (cf. Sect. 3.1.2) proffered explanation on the part of advocates of the dynamical approach to SR cannot be applied in the GR context. Thus, for the advocate of the dynamical approach, there are two brute facts in GR—two conspiracies, or 'miracles', which lack further explanation from within the theory—whereas in SR there is only one (see below):[26]

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

MR2: The Poincaré symmetries of the dynamical laws governing non-gravitational fields in the neighbourhood of any point $p \in M$ coincide (in the regime in which terms representing 'tidal gravitational forces' can be ignored) with the symmetries of the metric field in that neighbourhood.

[25] There exist significant difficulties regarding attempts to tell such a story of ontological reduction in GR; an obvious illustration can be found in the existence of vacuum solutions in the theory. This said, the question of whether an ontological excision of the metric field in GR is possible remains of philosophical and conceptual interest—particularly to advocates of the dynamical approach, for whom this would afford a means of bringing their approach to GR into line with their approach to SR.

[26] See Read et al. (2018, §5), where the terminology of 'miracles' was introduced, for further discussion.

There are two points to make here. First, note that **MR1** held also in SR: that all non-gravitational interactions are (locally) governed by Poincaré invariant dynamical laws is a *brute fact*—an *outset assumption*—in both theories, which (the advocate of the dynamical approach contends) admits of no further explanation from within each theory. Second, as I argue in Sect. 5.3, while an untenable form of the geometrical approach may purport to account for both **MR1** and **MR2**, *any acceptable form of the geometrical approach must also accept these two miracles of GR*. In this sense, the existence of these two miracles is independent of the dynamical/geometrical debate.

4 Qualified and Unqualified Explanations

In Sect. 5, I consider whether the advocate of the geometrical approach has received an unduly rough ride in the recent philosophical literature. Essentially, my answer will be affirmative, because an untenably strong form of the geometrical approach has constituted the target of e.g. Brown (2005), Brown and Pooley (2006), Brown and Read (2016), Read et al. (2018). In order to make these points, however, I must first distinguish between what I call *qualified* versus *unqualified* explanations:

- (*Qualified explanations*.) Consider one particular dynamical equation featuring coupling to a metric field—for example, the special relativistic Klein–Gordon equation (1), or the general relativistic Klein–Gordon equation (2). Then ask: might the metric field in the theory in question (η_{ab} in the case of **KGS**; g_{ab} in the case of **KGG**) feature in an *explanation* of the form (in particular, of the symmetries) of that dynamical equation, and of the behaviour of the matter field(s) (here φ) to which it is coupled? Call this the question of *qualified explanation*—for the concern here is with accounting for the form of one, *given* dynamical equation, and for the behaviour of the particular fields coupled in that equation.
- (*Unqualified explanations.*) Consider *all possible* dynamical equations consistent with a given theory, such as SR or GR.[27] Then ask: does the metric field in the theory in question (η_{ab} in the case of SR; g_{ab} in the case of GR) explain the form (in particular, the symmetries) of *all those possible equations* consistent with the theory, and (in a certain particular way to be articulated) the behaviour of all possible matter fields, such that assumptions made in the formulation of the theory regarding the form of those equations and the behaviour of matter fields (e.g., that massless particles in GR traverse null geodesics) are, ultimately, redundant? For example, can η_{ab} explain the fact that all dynamical laws

[27] There is some ambiguity regarding what is meant by a 'theory' here. To be clear, by 'theory' is meant here a theoretical *framework* such as that for SR or GR as presented in Sect. 2.3, rather than *specific* theories within those frameworks, such as **KGS** or **KGG**.

governing matter fields in SR are Poincaré invariant, or can g_{ab} in GR explain the SEP? Call this the question of *unqualified explanation*.

5 The Geometrical Approach

In this section, the above distinction between qualified and unqualified explanations is brought to bear on the question of whether there exists any viable form of the geometrical approach. My answer will be the following: while the form of the geometrical approach considered in e.g. Brown (2005), Brown and Pooley (2006), Brown and Read (2016), Read et al. (2018) is *not* viable, there exists a weaker version of the approach, which *can* be defended.

The section proceeds as follows. In Sect. 5.1, I distinguish between these two versions of the geometrical approach, before exploring the different accounts they give regarding the role of the metric field in explanations of the coincidence of (local) metric and dynamical symmetries, and of the behaviour of matter fields to which they couple, in both SR (Sect. 5.1.1) and GR (Sect. 5.1.2). In Sect. 5.2, I explore some further consequences of what I take to be the more defensible version of the geometrical approach. In Sect. 5.3, I demonstrate that this version of the geometrical approach does *not* account for **MR1** and **MR2**.

5.1 Two Geometrical Approaches

The geometrical approach, in both SR (Sect. 3.1.1) and GR (Sect. 3.2.1), may be understood in (at least) two different ways. Drawing upon the distinction presented in Sect. 4, the versions of the approach that I consider in this paper are dubbed the *qualified* versus *unqualified* geometrical approaches:

- (*Qualified geometrical approach.*) Consider a particular dynamical equation governing the behaviour of a particular set of non-gravitational fields $\Phi_1, \ldots, \Phi_n$. Insofar as that equation features coupling to a metric field (as in e.g. (1) in **KGS**, or (2) in **KGG**), that metric field may contribute to an explanation of the symmetries of that dynamical equation, and of the dynamical behaviour of those $\Phi_1, \ldots, \Phi_n$ fields.
- (*Unqualified geometrical approach.*) Consider the metric field associated with a particular theory (for example, η_{ab} in SR, or g_{ab} in GR). That metric field constrains the form of all possible dynamical laws for non-gravitational fields consistent with that theory, such that assumptions about (local) dynamical symmetries are redundant in the formulation of the theory, and such that certain facts about the behaviour of matter fields are fixed.

In the following, I abbreviate 'the qualified geometrical approach' to **QGA**, and 'the unqualified geometrical approach' to **UGA**. On **QGA**, a particular metric field coupling to a particular set of non-gravitational fields in a particular dynamical equation may be understood to contribute to a qualified explanation (in the sense in Sect. 4) of the symmetries of that dynamical equation, and of the dynamical behaviour of those non-gravitational fields. On **UGA**, a particular metric field is taken to explain the symmetries of *all possible* dynamical equations in a given theory, and to fix certain facts about the behaviour of all possible matter fields, such that we need not, in fact, make any assumptions regarding dynamical symmetries, or about those dynamical facts, in that theory. (Importantly, I take both **QGA** and **UGA** to maintain the ontological autonomy of the metric field in both SR and GR.) It is principally **UGA** which is attacked in Brown (2005), Brown and Pooley (2006), Brown and Read (2016), Read et al. (2018), and it is this version of the geometrical approach which is (I maintain) untenable.

5.1.1 Special Relativity

The reasons why **UGA** is untenable are similar in both the SR and GR cases; I begin with the former. The worry regarding **UGA** is put clearly by Brown and Pooley (2006, p. 84):

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

The point here is that it is consistent to have dynamical laws for non-gravitational fields in a theory featuring a Minkowski metric field η_{ab}, which nevertheless *do not* manifest the Poincaré symmetries of that metric field. As a concrete example, consider a modified version of **KGS**—call it **LAS**—KPMs of which are quadruples $\langle M, \eta_{ab}, \delta_{ab}, \varphi \rangle$, where δ_{ab} is a four-dimensional fixed *Euclidean* metric field,[28] and DPMs of which are picked out by the four-dimensional *Laplace* equation (hence my chosen nomenclature),[29]

$$\delta_{ab} \nabla^a \nabla^b \varphi = 0. \tag{10}$$

The dynamical symmetries of (10) do *not* include the Poincaré transformations (as for (1)); rather, they include the *Euclidean* transformations: those affine transformations the linear transformation matrix of which satisfies (cf. (5))

$$M_{\mu'}{}^{\mu} M_{\nu'}{}^{\nu} \delta_{\mu\nu} = \delta_{\mu'\nu'}. \tag{11}$$

---

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

[29] Note that (10) is simply (1), with η_{ab} replaced by δ_{ab}; in making this move, the dynamical equation becomes an *elliptic*, rather than *hyperbolic*, partial differential equation.

LAS illustrates that a theory's featuring a certain metric field in its KPMs is *insufficient* for that theory's dynamical equations for non-gravitational fields to manifest the symmetries of that metric field, or for that metric field to play any constraining role in the dynamics of the matter fields in that theory, for those non-gravitational fields may couple to *other* fields (in this case, δ_{ab}), such that metric symmetries and dynamical symmetries do *not* coincide, and such that the matter fields manifest *other* dynamical behaviour (than that which they would manifest if they were coupled to the metric field under consideration, here η_{ab}). Of course, one may wish to exclude coupling to such other fields; however, note that we then return to the situation in which the dynamical equations for matter fields manifesting certain symmetries, and yielding certain behaviour for those matter fields (e.g., that massless particles propagate on null geodesics), is an input *assumption*—it does not follow from (e.g.) η_{ab} alone.

On the other hand, **QGA** faces no such problems, for in this case the concern is not with generic, unqualified claims, but rather with the form of *one particular* dynamical equation and with the dynamical behaviour of the matter fields coupled in that equation. Why is (1) Poincaré invariant? Because it features coupling to the η_{ab} field—cf. Sect. 2.2. Why is (10) Euclidean invariant? Because it features coupling to δ_{ab}. Changing η_{ab} in (1) to δ_{ab} in (10) changes the behaviour of φ accordingly (after all, it is now governed by a different dynamical equation)—and it is very plausible to regard this as constituting a legitimate (if partial, for other factors may also be relevant to the dynamics of the field in question) *explanation* of the behaviour of φ. Thus, I take it that, in SR (and indeed, in the context of theories with fixed metric structure more generally), it is *incorrect* to regard as viable the explanation for the dynamical behaviour of matter proffered on the part of advocates of **UGA**, but *correct* to so regard the explanations proffered on the part of advocates of **QGA**. I discuss **QGA** further in Sect. 5.2.

5.1.2 General Relativity

Similar points to those made above apply in the case of GR. According to advocates of **UGA**, the metric field g_{ab} in GR accounts for the local behaviour of all non-gravitational fields, such that the assumption of the SEP in the presentation of general relativistic theories in Sect. 2.3.2 is redundant, and such that matter fields must exhibit certain behaviour (e.g., such that test particles propagate on null geodesics). However, against such a claim, problem cases may also be identified.

In parallel with our introduction of **LAS** in Sect. 5.1.1, I present here one such theory: the *Jacobson–Mattingly theory* (introduced in e.g. Carroll and Lim 2004; Jacobson and Mattingly 2001[30]), in which the action for a coupled Einstein–Maxwell system is augmented with an additional term (via a Lagrange multiplier

[30]In fact, the version of the Jacobson–Mattingly theory discussed here is a special case of that presented in Carroll and Lim (2004), Jacobson and Mattingly (2001).

field λ), imposing (as a field equation, via variation with respect to λ) that the vector potential A^a be locally timelike:[31]

$$S_{\mathrm{JM}}\left[g_{ab}, A^a, \lambda\right] = \int d^4x\sqrt{-g}\left(R - \frac{1}{4}F^{ab}F_{ab} + \lambda\left(g_{ab}A^aA^b - 1\right)\right). \quad (12)$$

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

As with **LAS** in the case of SR, such cases appear to find their mark against **UGA**, for in the Jacobson–Mattingly theory, metric symmetries manifestly do *not* coincide with dynamical symmetries—so how could g_{ab} be constraining the local form of dynamical equations in this strong sense? On the other hand, **QGA** again does not appear to face such problems. For example, consider (2)—as in the SR context, it is perfectly reasonable to claim that the coupling in this equation of φ to g_{ab} offers an explanation of the dynamical behaviour of φ; moreover, the fact that no generic, unqualified claim is made regarding possible form of dynamical equations for non-gravitational fields means that cases such as the Jacobson–Mattingly theory do not find their mark against **QGA** (for further discussion, see Sect. 5.2.2).

5.2 *The Qualified Geometrical Approach*

I have argued that **QGA** is a defensible version of the geometrical approach, whereas **UGA** is not. In this subsection, I explore some further consequences of **QGA**. Specifically, I consider in Sect. 5.2.1 the sense in which the metric field in a given theory *may*, in fact, be understood to account for the form of *all* dynamical laws in that theory. In Sect. 5.2.2, I consider whether an account of the dynamical behaviour of matter in terms of metric structure is available on **QGA**, even in problematic cases such as those described above, in which (local) metric symmetries do not coincide with (local) dynamical symmetries.[32] I close in Sect. 5.2.3 by drawing a more fine-grained distinction within **QGA**.

[31] The first term is the Einstein–Hilbert action; F_{ab} is the Faraday tensor associated with A^a. In this paper, I take it that in GR (or, as here, the Jacobson–Mattingly theory) a vector ξ^a at a point is *timelike* just in case $g_{ab}\xi^a\xi^b < 0$.

[32] Strictly, I will have to generalise the notion of a 'metric symmetry' in Sect. 5.2.2, to account for the examples given in that section. This, however, will be of no consequence.

5.2.1 Univocal Explanation

In both SR and GR, there is a sense in which, on **QGA**, the metric field *can* explain the form of all dynamical laws in the theory—*once the restriction to a certain form of dynamical equations is made*. For example, *given* the restriction in SR to dynamical equations for non-gravitational fields which take a Poincaré invariant form, we may write all such equations in coordinate-free notion featuring coupling to η_{ab}[33]—in which case, η_{ab} may feature in explanations of the dynamical behaviour of the matter fields under consideration. This does not explain the *initial* restriction to Poincaré invariant dynamical laws for non-gravitational fields, but it *does* mean that η_{ab} may feature in explanations for the behaviour of all matter fields, *once such an assumption is made*. Similarly in GR, the metric field g_{ab} may not be able to account for the *initial* restriction to dynamical equations for matter fields obeying the SEP, but it *may* feature in explanations of the form of all dynamical laws for non-gravitational fields in GR, *once this assumption is made*—for in making this assumption, it is natural to consider dynamical equations in which matter fields are coupled to this very g_{ab} field.[34]

5.2.2 Partial Explanation

A further subtlety regarding **QGA** pertains to the issue of *partial* explanation. I make the following claim: even in the cases in which metric and dynamical symmetries do not coincide, the metric field *may* feature in explanations of the dynamical behaviour of matter, on **QGA**.[35] To see this, it is useful to consider three sub-cases: (1) situations in which dynamical symmetries form a proper subset of metric symmetries; (2) situations in which dynamical symmetries form a proper superset of metric symmetries; (3) cases where dynamical symmetries partially overlap with metric symmetries.

[33]Cf. Brown and Read (2016, §5).

[34]It is worth making two related points here. (1): Technically, such coupling is not essential, for we might instead couple to e.g. a fixed Minkowski metric field η_{ab}, or to a generic Lorentzian metric field which satisfies not the Einstein field equations, but some other set of dynamical equations. In the cases in which all dynamical laws feature coupling to g_{ab}, however, this metric field may feature in explanations of the form of all these laws. (2): One need not make the assumption that all dynamical laws manifest certain (local) symmetries so explicitly—one might instead make assumptions of (e.g.) *universal coupling* of the metric field to matter fields in *all* dynamical equations for the latter; this may, then, *entail* the relevant facts about the symmetries of those laws. This, indeed, appears to be Maudlin's stance, when he writes that 'the fundamental requirement of a relativistic theory is that the physical laws should be specifiable using only the relativistic space-time geometry. For Special Relativity, this means in particular Minkowski space-time.' Maudlin (2012, p. 117) The point here is that, on **QGA**, one may appeal to the metric field in giving certain generic explanations of the behaviour of matter fields in a certain restricted class of models of the theory—but the metric field itself does not account for those restrictions.

[35]I am grateful to Oliver Pooley for impressing this point upon me.

In order to discuss each of these cases, it is useful to introduce here three versions of Newtonian gravitation theory (NGT). First, let a *Leibnizian structure* be a triple $\langle M, t_{ab}, h^{ab} \rangle$, where M is a four-dimensional differentiable manifold; t_{ab} is a fixed temporal 'metric' field on M of signature $(1, 0, 0, 0)$; and h^{ab} is a fixed spatial (inverse) 'metric' field on M of signature (0, 1, 1, 1).[36] The t_{ab} and h^{ab} fields are orthogonal, so that

$$h^{ab} t_{bc} = 0; \tag{13}$$

furthermore, I restrict in this paper to structures (Leibnizian or otherwise; see below) which are *temporally orientable*, so that there exists a continuous (globally defined) one-form t_a that satisfies the decomposition condition $t_{ab} = t_a t_b$ at every point (Malament, 2012, p. 251).

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

$$\nabla_a t_{bc} = 0, \tag{14}$$

$$\nabla_a h^{bc} = 0. \tag{15}$$

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

$$t_{ab} \sigma^b \neq 0. \tag{16}$$

Since none of Leibnizian, Galilean, or Newtonian structures are themselves metric fields, the notion of a metric symmetry cannot be applied in these cases.[37] However, the relevant notion easily generalises to the structures now under consideration: I say that a coordinate transformation is a *structure symmetry* just in case the structure under consideration is invariant under that transformation. Applying such a notion to Leibnizian, Galilean, and Newtonian structures, one finds that their associated structures symmetries are given by (no surprise!) the Leibniz, Galilean, and Newton groups.[38]

With these three structures in hand, we can consider three different theories—viz., Newtonian gravitation theory set in each of these three structures. Consider first Newtonian mechanics set in a Galilean structure.[39] KPMs of this theory are

[36]Scare quotes are included on 'metric' here, for strictly neither t_{ab} nor h^{ab} satisfies the metric non-degeneracy condition—cf. Malament (2012, §4.1).

[37]For details regarding Leibnizian, Galilean, and Newtonian structures, see Earman (1989, ch. 2).

[38]The exact mathematical forms of these groups are not relevant for our purposes—see Pooley (2013, §3.1) for details.

[39]A Galilean structure is traditionally considered to be the 'most appropriate' spacetime setting for NGT, for in this case structure symmetries and dynamical symmetries (are claimed to) coincide,

tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \varphi, \rho\rangle$, where φ and ρ are real scalar fields on M, which will be taken to represent the gravitational potential and matter density, respectively. DPMs of this theory are picked out by the field equations[40]

$$R^a{}_{bcd} = 0, \tag{17}$$

$$h^{ab}\nabla_a\nabla_b\varphi = 4\pi\rho. \tag{18}$$

Equation (17) imposes flatness of ∇_a; (18) is the Newton–Poisson equation. Finally, the gravitational force on a point (test) particle of mass m is given by $-mh^{ab}\nabla_b\varphi$; it follows from Newton's second law that, if this particle is subject to no forces except gravity, and given that it has four-velocity ξ^a, then it satisfies

$$-\nabla^a\varphi = \xi^b\nabla_b\xi^a. \tag{19}$$

Note that all elements of the Galilean structure feature in these dynamical equations; one can use this structure to offer a qualified explanation (in the sense of Sect. 4) of the form of these dynamical laws.

Newtonian mechanics set in Galilean spacetime is a case in which structure symmetries coincide with dynamical symmetries.[41] Now consider a more nuanced case, in which dynamical symmetries constitute a proper subset of structure symmetries. One illustration of this is Newtonian mechanics set in a Leibnizian structure. KPMs of this 'theory' are tuples $\langle M, t_{ab}, h^{ab}, \varphi, \rho\rangle$ with $\langle M, t_{ab}, h^{ab}\rangle$ a Leibnizian structure, and φ and ρ defined as in the Galilean case; DPMs are (allegedly) picked out by (17)–(19). For the sake of argument granting that such a 'theory' is coherent,[42] we have a case in which dynamical symmetries are a proper subset of structure symmetries. What I contend here is that, in spite of the fact that structure symmetries and dynamical symmetries do not coincide, the fact that the Leibnizian structure still features in the DPMs of this theory means that it can still offer a *partial* (but not complete, since the laws also advert to other structure) explanation of the dynamical behaviour of matter in this case, in the qualified sense delineated in Sect. 4 above.

thereby satisfying Earman's 'adequacy conditions' on spacetime theories (see Earman 1989, §3.4). For recent philosophical discussion calling into question whether this orthodoxy is correct, see Dewar (2018), Knox (2014), Saunders (2013), Teh (2018), Wallace (2017), Weatherall (2016a, 2018); I do not discuss further such matters in this paper.

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

[41]Setting aside the issues indicated in footnote 39.

[42]Indeed, I here include scare quotes on the word 'theory', as there are good grounds to question whether such a 'theory' is really coherent, since it does not have sufficient structure in its KPMs to be able to write down the dynamical equations used to fix its DPMs—cf. Stein (1977, p. 6). (Belot puts the point pithily, when he accuses those working with such theories of 'arrant knavery' Belot 2000, p. 571; for further related discussion, cf. Dewar 2018, pp. 268–269.)

Next consider the case in which dynamical symmetries are a proper superset of structure symmetries.[43] An illustration of such a scenario is Newtonian mechanics set in a Newtonian structure. In this case, KPMs are tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho\rangle$, where $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a\rangle$ is a Newtonian structure (in which the integral curves of σ^a are taken to represent the worldlines of the persisting points of Newtonian absolute space), and φ and ρ are understood as above; DPMs of this theory are again picked out by (17)–(19). Though this theory is coherent,[44] as Earman states (Earman, 1989, §3.4), there is a sense in which it is nevertheless *malformed*, for the dynamical laws do not advert to all the Newtonian structure available in the KPMs of the theory (it is this which results in dynamical symmetries being a proper superset of structure symmetries). While I concur with Earman on this point, what I wish to register here is that, in this case, Newtonian structure *may* still be appealed to in explanations of the form of the dynamical laws governing matter fields—it is just that this structure has other, *redundant* explanatory machinery available to it (viz., the σ^a field).

Thus, on **QGA**, in the case in which dynamical symmetries are a subset of structure symmetries, the relevant structure (whether metric, or e.g. Leibnizian/Galilean/Newtonian) may feature in *partial* explanations of the dynamical behaviour of matter. In the case in which dynamical symmetries are a superset of structure symmetries, by contrast, the relevant structure may feature in *total* but *redundant* explanations of the dynamical behaviour of matter. Note that Jacobson–Mattingly theory of Sect. 5.1.2 instantiates the former case, in which dynamical symmetries are a subset of metric symmetries.[45]

Finally, consider the case in which dynamical symmetries partially overlap with structure symmetries—i.e., are neither a subset nor a superset of structure symmetries. One example of this is **LAS**, presented in Sect. 5.1.1. In this case, dynamical symmetries include the Euclidean transformations; symmetries of the Minkowski metric field η_{ab} are the Poincaré transformations. The intersection of the Euclidean and Poincaré groups is the group of translations and spatial rotations (cf. Read et al. 2018, §B); therefore, the corresponding degrees of freedom associated with the η_{ab} field may still be used to account for *these* dynamical symmetries, in this case. (Though of course, an obvious question arises: why not instead appeal to δ_{ab} when giving this kind of qualified explanation of dynamical symmetries in this case?)

[43] On this possibility, cf. Pooley's discussion at Pooley (2013, p. 94).

[44] At least on **QGA**—it is questionable whether this theory is coherent on the dynamical approach, according to which (as discussed above) metric/structure symmetries in theories with fixed metric/structure (such as both SR and NGT) *just are* dynamical symmetries. Cf. Brown and Read (2020, §3.1).

[45] Though in this case the theory is coherent, in a way that arguably NGT set in a Leibnizian structure is not—cf. footnote 42.

5.2.3 Confident and Cautious Qualified Approaches

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

Different possible answers to this question distinguish two sub-views within **QGA**. On the one hand, one might maintain that, when the metric/structure features in a qualified explanation of the symmetries of the dynamical laws governing matter fields in the above sense, there always exist physical rods and clocks built from matter fields which survey that metric/structure. Call this view *confident* ***QGA***.[46] On the other hand, one might reject the claim that, when the metric/structure features in a qualified explanation of the symmetries of the dynamical laws governing matter fields in the above sense, there always exist physical rods and clocks built from matter fields which survey that metric/structure. Call this view *cautious* ***QGA***.

Clearly, in order to call into question confident **QGA**, it suffices to present a single problem case. In fact, there exist several such cases; here I mention two. First, Pitts (2019) presents the example of *universally coupled massive scalar gravity*. In such theories, there exist two Lorentzian metric fields: a dynamical field g_{ab}, and a fixed Minkowski metric field η_{ab}; the Lagrangian includes the following graviton mass piece:[47]

$$\mathcal{L}_{\text{mass}} = \frac{m^2}{64\pi G}\left[\frac{\sqrt{-g}}{w-1} + \frac{\sqrt{-g}^{\,w}\sqrt{-\eta}^{\,1-w}}{w\,(1-w)} - \frac{\sqrt{-\eta}}{w}\right]. \tag{20}$$

(Here, w is a free parameter, which may be fixed to yield specific theories.) The important point to note about such theories is put clearly by Pitts: ‘Massive scalar gravity lacks Minkowskian behaviour of rods and clocks, though it has the Minkowski metric (among other things) and the Poincaré symmetry group. . . . [T]he chronogeometrically observable conformally flat metric $g_{ab} = \hat{\eta}_{ab}(-g)^{1/4}$ isn’t

[46] Arguably, Maudlin falls into this camp, for he both (a) speaks of restricting dynamical equations in SR to those which couple universally to η_{ab}, thereby placing him in **QGA** (cf. footnote 34); and (b) argues that, in any model of SR, there exists a clock which satisfies the *clock hypothesis*, and thereby (by definition) correctly reads off intervals along its worldline as given by the metric field (cf. Maudlin 2012, ch. 5). There are good reasons to doubt (b)—cf. Menon et al. (2018), discussed further below.

[47] For the full details, see Pitts (2019).

clearly the One True Geometry.'[48] (Pitts, 2019, p. 6). Thus, theories of this kind appear to pose problems for confident **QGA**, for rods and clocks generally *do not* survey η_{ab}, in spite of the fact that this field couples to the matter fields in the theory, and so may feature in a qualified explanation of their symmetries.

As a second example, the authors of Menon et al. (2018) demonstrate, drawing upon recent work by Asenjo and Hojman (2017), that there should be no expectation that physical rods and clocks (such as light clocks) correctly survey the metric field g_{ab} of GR in particular solutions of this theory—namely in rotating solutions, such as the Gödel and Kerr solutions. The reasons are subtle, but essentially involve the fact that physical propagating media, such as light waves, do not travel at a fixed speed in such solutions, but rather manifest spacetime location-dependent propagation speeds. The central point here is a simple one: there is again reason to doubt confident **QGA**, for in these cases one has dynamical equations governing matter fields which feature coupling to g_{ab}, so that this metric field may feature in a qualified explanation of the symmetries of these equations and the behaviour of matter fields; nevertheless, rods and clocks do not survey this metric field, so that the chronogeometric significance of this field is questionable.

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

5.3 Two Miracles, Reprise

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

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

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

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

In the GR context, **QGA** also faces both **MR1** and **MR2**—for exactly the reasons delineated in Sect. 3.2.2. Given this, a new question arises: given that both the dynamical approach and **QGA** agree in the GR context that the g_{ab} field cannot be ontologically reduced to matter fields, and that both **MR1** and **MR2** hold in that context, is there really such a difference between the views, in this case? Absent the story of ontological reduction, there appears to be very little between the views. In light of this, I make the following claim: *While the dynamical approach and **QGA** are distinct in the context of theories with fixed metric structure such as SR (for they make different ontological claims regarding this fixed structure), they are not distinct in the context of theories with dynamical metric structure, such as GR.*[49]

6 The Geodesic Principle

So far, I have: (a) clarified the distinction between the dynamical and geometrical approaches—the latter itself coming in two distinct varieties: **UGA** and **QGA**; (b) argued that while **QGA** is viable, **UGA** is not; (c) demonstrated that **MR1** and **MR2** hold both on the dynamical approach and on **QGA**; (d) argued that there is no difference between **QGA** and the dynamical approach in the context of GR. In this section, I consider the connections between this work, and recent and important results on the *geodesic principle*. I also reflect upon work by Knox (2017, 2013, 2014) and Weatherall (2017, §6) pertinent to the themes of this paper.

I begin with the geodesic principle. Contemporary work on this result stems largely from a 1975 theorem of Geroch and Jang (1975). Though more sophisticated extensions of this result now exist (in particular, see Ehlers and Geroch 2004;

[49] Again, I am grateful to Oliver Pooley for impressing this point upon me. In this regard, cf. Pooley (2013, p. 63), where Pooley writes, 'What, then, is at stake between the metric-reifying relationalist and the traditional substantivalist? Both parties accept the existence of a substantival entity, whose structural properties are characterised mathematically by a pseudo-Riemannian metric field and whose connection to the behaviour of material rods and clocks depends on, *inter alia*, the truth of the strong equivalence principle. It is hard to resist the suspicion that this corner of the debate is becoming merely terminological.'

Geroch and Weatherall 2018), I focus for the time being upon the Geroch–Jang theorem itself; this reads as follows:[50]

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

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

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

The Geroch–Jang theorem makes precise the essence of the geodesic principle: that small bodies move on geodesics. In Weatherall (2017, §6), Weatherall draws a number of philosophical lessons regarding geodesic theorems such as the above (and its more sophisticated successors), which he takes to be consonant with the dynamical approach; it is to these putative lessons that I now turn.[51] Begin with Weatherall's summary of the import of results such as the Geroch–Jang theorem:[52]

> [E]stablishing that small bodies respect the inertial structure encoded by a given derivative operator ∇_a requires one to establish that the T^{ab} field associated with matter is divergence-free, or "conserved", with respect to ∇_a (Weatherall, 2017, p. 36).[53]

Weatherall takes the fact that T^{ab} is conserved with respect to a *specific* derivative operator ∇_a to deliver a connection between satisfaction of the geodesic principle and spacetime geometry—with this being particularly apparent if that derivative

[50] Here, I use the notation of Weatherall (2017, p. 6).

[51] For Brown's own discussion of the geodesic principle, see Brown (2005, §9.3). With Brown's central contention—that geodesic motion of small bodies in GR is a consequence of the Einstein field equations, and is therefore automatic in GR, in a way that it is not in antecedent theories ('It is no longer a miracle.' Brown (2005, p. 163))—Weatherall is in disagreement, for (a) geodesic motion is, in fact, independent of the Einstein field equations; (b) similar results can be derived in other theories, e.g. NGT, and Newton–Cartan theory. (For the details of Newton–Cartan theory, in which the gravitational potential φ of NCT is absorbed into a (curved) derivative operator, see Malament 2012, ch. 4.) For Weatherall's work on the geodesic principle, see Weatherall (2017, 2011c,b, 2012, 2017); I am in agreement with him on these matters. Also worthy of mention in this regard are remarks in a similar vein to (a) made by Pooley (2013, p. 543); and an earlier paper of Malament (2012), in which it is pointed out (*pace* Brown) that geodesic motion in GR follows *only* on the assumption of the strengthened dominant energy condition.

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

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

operator ∇_a is that which is compatible with some metric field: (Weatherall 2017, p. 38)

> From this perspective it is also fair to say that, as Brown argues in *Physical Relativity*, spacetime structures such as the metric may be viewed as "a codification of certain key aspects of the behaviour of particles and fields" (p. 142), at least as regards the link between free, small-body motion and the privileged class of curves picked out by a metric and/or derivative operator.

Though I am in agreement with Weatherall as far as the above statements go, there remains more to be said here, on two fronts. First, though it is true that some connection between matter fields and geometry is forged insofar as the stress-energy tensor associated with these fields is conserved with respect to a specific derivative operator, and moreover insofar as that matter thereby follows geodesics of that derivative operator, in accordance with the Geroch–Jang theorem (or its extensions), thus far the connection proceeds in terms of the motion of *small bodies alone*. To move from such results regarding the geodesic motion of small bodies, to the behaviour of matter fields *tout court*, is in effect to demand that the local symmetry properties of all matter fields be derivable from such geodesic motions; that is, it is, in effect, to demand a proof of a result akin to *Schiff's conjecture*.[54] Only in that case could something like the SEP be delivered by this work on the geodesic principle.[55]

Second, it is important to be clear that this work does not provide a resolution to **MR1**. Even supposing that a connection is forged between geodesic motion and

[54]In the words of Thorne et al., '*Schiff's conjecture states that any complete and self-consistent gravitation theory that obeys [the weak equivalence principle] must also, unavoidably, obey [the strong equivalence principle]*' (emphasis in original) (Thorne et al., 1973, p. 3575). In turn, the weak equivalence principle is defined as follows: '*If an uncharged test body is placed at an initial event in spacetime, and is given an initial velocity there, then its subsequent worldline will be independent of its internal structure and composition*' (emphasis in original) (Thorne et al., 1973, p. 3571); the strong equivalence principle is defined as: '(i) [The weak equivalence principle] is valid, and (ii) the outcome of any local test experiment—gravitational or nongravitational—is independent of where and when in the universe it is performed, and independent of the velocity of the (freely falling) apparatus' (Thorne et al., 1973, p. 3572). For the original presentation of Schiff's conjecture, see Schiff (1960, p. 343); for ensuing discussion and attempted proofs of restricted versions of the conjecture, see Coley (1982), Lightman and Lee (1973), Ni (1977), Thorne et al. (1973). Clearly, the version of Schiff's conjecture under consideration in this paper is different to that above—the gap to be bridged here is between the geodesic motions of small bodies, and the symmetries of matter fields *tout court*.

[55]Geroch and Weatherall demonstrate in Geroch and Weatherall (2018) that source-free Maxwell fields 'track' null geodesics—a new result. Since the geodesic theorems demonstrate that massive matter moves on timelike geodesics, this gives access to both conformal and projective structure, respectively. One might think, therefore, that one may appeal to the Ehlers–Pirani–Schild result Ehlers et al. (1972) (itself a generalisation of Weyl's theorem—cf. Weyl 1921), that (subject to extra constraints) conformal and projective structure fixes metric structure, to strengthen the connection between these geodesic theorems and geometry. While such results do indeed yield a further sense in which local geometry may be inferred from geodesic motions, they continue to leave unbridged the gap between the geodesic motions of small bodies, and the local dynamics of matter *tout court*. That is, Schiff's conjecture remains unproven, in general.

the local behaviour of matter fields more generally (*à la* Schiff's conjecture), that the mystery of **MR1** remains can be demonstrated through asking the following question: why should all matter fields have associated stress-energy tensors, the divergences of which vanish with respect to the *same* derivative operator? If this were not the case, then it need not be the case that all matter fields survey the same 'practical geometry', in the manner explicated by Weatherall. Though it is true that, as Weatherall observes (Weatherall, 2017, p. 11), the Einstein field equations tell us (via the contracted Bianchi identity) that the covariant divergence of the *total* stress-energy content of any particular solution of GR vanishes, this is (again, as Weatherall observes—see Weatherall 2017, p. 12) insufficient to infer that the divergences of the stress-energy tensors associated with all *individual* matter fields vanish with respect to the same derivative operator. Thus, these results on the geodesic theorem do not place sufficient restrictions on the behaviour of even small bodies built from different matter fields in order to resolve **MR1**.

The situation regarding the bearing of these results upon **MR2** is more nuanced. Suppose that if the dynamical laws governing matter fields all manifest the same symmetries, then the stress-energy tensors associated with such matter fields (which satisfy the strengthened dominant energy condition, and the other conditions of the Geroch–Jang theorem and its generalisations) have covariant divergences which vanish with respect to the same derivative operator. Now suppose that the dynamical laws governing matter fields all manifest the same symmetries. Then (by the above), the stress-energy tensors associated with such matter fields have covariant divergences which vanish with respect to the same derivative operator. Then, divergence of the total stress-energy tensor (being a sum of the stress-energy tensors associated with the individual matter fields) with respect to this same derivative operator will also vanish; so, via the Einstein field equations, the left-hand side of the field equations will also have vanishing divergence with respect to this derivative operator—implying that the derivative operator is compatible with the metric field appearing in the Einstein tensor. In that case, small bodies built from all matter fields 'track' geodesics of a derivative operator associated with the Lorentzian metric field appearing in the Einstein field equations. In turn, one expects that in such a case the symmetries of the dynamical laws governing matter fields, and of this metric field, coincide, thereby delivering **MR2**. Of course, this reasoning is heuristic—but renders it *prima facie* plausible that these results regarding the geodesic principle may have application in resolving **MR2**.

In any case, let us now set aside these considerations regarding **MR1** and **MR2**, and focus upon Weatherall's general morals drawn in Weatherall (2017, §6). Consider the following passage:

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

> (all) matter in such a way that all inner products are taken relative to that metric and all derivatives are taken relative to that derivative operator. This is the physical content of the claim that there are facts about distances, angles, and duration: physical processes occur in such a way that changes in a quantity at a time depend on the state of that quantity and those facts about distances, angles, and duration. And so, one is left with the conclusion that spacetime structures codify certain facts about the behavior of matter because the dynamics of (all) matter is adapted to those spacetime structures, which is just another way of saying that spacetime 'has' that geometry (Weatherall, 2017, pp. 39–40).

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

Second, nothing in this passage commits Weatherall either to the dynamical approach, or to **QGA**. Insofar as Weatherall takes e.g. NGT set in a Newtonian structure to be a coherent theory, there is perhaps some reason to take him to favour the latter, for recall that the coherence of this theory is questionable on the dynamical approach—cf. Brown and Read (2020, §3.1).[56] Even in this case, however, one might take Weatherall's anticipated assessment that this theory is 'theoretically equivalent' (in a technical, category-theoretic sense—cf. Weatherall 2016; Weatherall 2016b, 2018) to NGT set in a Galilean structure, combined with an implicit commitment to such theoretical equivalence being sufficient for physical equivalence, to indicate that he does *not* consider such to be the case—meaning that perhaps he should be regarded as siding with advocates of the dynamical approach after all.[57]

Third, Weatherall's views as expressed in the above passage are very consonant with the 'spacetime functionalism' of Knox (2013, 2014, 2017), according to which 'the spacetime role is played by whatever defines a structure of local inertial frames' (Knox, 2017, p. 22) (cf. Sect. 1). To see this, some details regarding this programme of Knox must be recalled. Note first that, in GR, the chronogeometricity of the

[56]In more detail, recall from footnote 44 that, on the dynamical approach, metric/structure symmetries in theories with fixed metric/structure *just are* dynamical symmetries—so how could it be the case that there exists a theory in which such symmetries do not coincide?

[57]I concede that it is somewhat strained to seek to read Weatherall as an advocate of the dynamical approach; a reading on which he endorses something like **QGA** is more natural. Nevertheless, it is at least worth noting that advocation of the dynamical approach is *consistent* with Weatherall's writings. (Moreover—and interestingly—Weatherall has questioned in personal communication whether fixed metric structure, such as the Minkowski metric field of SR, should be regarded as being ontologically autonomous—in which case, his views are arguably closer to the dynamical approach than one might initially think. Whether, however, it is best to read Weatherall as endorsing the dynamical approach versus e.g. the version of the geometrical approach due to Janssen (2009), Balashov and Janssen (2003), Janssen (2002), in which the ontological autonomy of the metric field in e.g. SR is denied, remains unclear absent further work. Since the issues here are subtle, and it would take significant work to do justice to Janssen, these matters will have to wait for a future piece.)

metric field precisely guarantees that this field be considered spatiotemporal, in Knox's sense. The reason is that, locally, the symmetries of the dynamical metric field coincide with those of the dynamical equations governing matter fields; in any frame in which these dynamical equations take their simplest form, the metric field itself takes the form diag $(-1, 1, 1, 1)$. Thus, the metric field picks out a structure of local inertial frames—if one characterises such frames as those in which dynamical equations for non-gravitational fields take their simplest form (cf. Knox 2013, §2).

Now recall that, for Weatherall, 'what one means, or at least what one should mean, by the claim that spacetime has some geometry, represented by a given metric (or metrics) and derivative operator, is precisely that one can express the dynamics of (all) matter in such a way that all inner products are taken relative to that metric and all derivatives are taken relative to that derivative operator' (Weatherall, 2016a, p. 40). But, so coupling the dynamical equations governing matter fields will in general ensure that those equations have certain local symmetry properties—as, for instance, our discussion of **KGG** illustrated. In particular, it will in general ensure that metric symmetries coincide (locally) with dynamical symmetries—that is, it will ensure that the metric field qualifies as spatiotemporal, on Knox's programme.[58] Thus, when Weatherall states that such coupling is sufficient for 'spacetime to have some geometry', I take it that he is endorsing a view very much akin to Knox's spacetime functionalism.[59]

7 Conclusions

In the context of SR (and of theories with a fixed metric/structure[60] more generally), advocates of the dynamical approach maintain that such a metric/structure is ontologically reducible to (symmetries of the dynamical laws governing) non-gravitational fields. By contrast, in the context of GR (and of theories with a dynamical metric/structure more generally), no such claim is made on the part of advocates of the dynamical approach. As a result of this, the dynamical approach arguably collapses into **QGA** in the GR context. While the dynamical approach is distinct from **UGA** in both SR and GR, there are good reasons to doubt the plausibility of **UGA**.

On **QGA**, we can appeal to the metric field of e.g. SR or GR to explain certain universal facts about the dynamics of matter fields—but only once further restrictions on the allowed class of models under consideration are imposed (for

[58] This coupling will ensure that a necessary condition on the metric field's having chronogeometric significance is satisfied—cf. Sect. 3.

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

[60] 'Structure' construed here in the sense of Sect. 5.2.2.

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

Weatherall may be understood as embracing Knox's spacetime functionalism, alongside either the dynamical approach, or **QGA**. Since both the dynamical approach and **QGA** are defensible, this is unproblematic. Indeed, arguably the geometrical approach has been written off too quickly by advocates of the dynamical view, as a result of a lack of appreciation of the viability of **QGA**. While I incline to the dynamical view—essentially on grounds of ontological parsimony in theories such as SR—I hope this paper may be of some value in demonstrating that the views of essentially all parties in this debate do not stand in such a state of conflict as one may *prima facie* be inclined to think.

Acknowledgments I am grateful to Erik Curiel, Patrick Dürr, Niels Linnemann, Tushar Menon, Oliver Pooley, and Jim Weatherall for invaluable comments on earlier drafts of this paper; to Harvey Brown, Jeremy Butterfield, Dennis Lehmkuhl, and Brian Pitts for fruitful discussions on the dynamical approach; and to Jim Weatherall (again) for asking a pertinent question which motivated my writing this paper. I am also indebted to audiences in Bern, Bonn, Cambridge, Exeter, Kraków, and London for useful feedback. My D.Phil. studies were supported by an AHRC grant at the University of Oxford; a senior scholarship at Hertford College, Oxford; and a fellowship with the Space and Time After Quantum Gravity project at the University at Illinois at Chicago and the University of Geneva—the latter funded by the John Templeton Foundation.

References

Pablo Acuña, "Minkowski Spacetime and Lorentz Invariance: The Cart and the Horse or Two Sides of a Single Coin?", Studies in History and Philosophy of Modern Physics 55, pp. 1–12, 2016.

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

Felipe A. Asenjo and Sergio A. Hojman, "Do Electromagnetic Waves Always Propagate Along Null Geodesics?", Classical and Quantum Gravity 34, 205011, 2017.

Yuri Balashov and Michel Janssen, "Presentism and Relativity", British Journal for the Philosophy of Science 54(2), pp. 327–346, 2003.

Gordon Belot, "Geometry and Motion", British Journal for the Philosophy of Science 51, pp. 561–595, 2000.

Harvey R. Brown, *Physical Relativity: Spacetime Structure from a Dynamical Perspective*, Oxford: Oxford University Press, 2005.

Harvey R. Brown and Oliver Pooley, "The Origins of the Spacetime Metric: Bell's Lorentzian Pedagogy and its Significance in General Relativity", in Craig Callender and Nick Huggett (eds.), *Physics Meets Philosophy at the Plank Scale*, Cambridge: Cambridge University Press, 2001.

Harvey R. Brown and Oliver Pooley, "Minkowski Space-Time: A Glorious Non-Entity", in Dennis Dieks (ed.), *The Ontology of Spacetime*, Elsevier, 2006.

Harvey R. Brown and James Read, "Clarifying Possible Misconceptions in the Foundations of General Relativity", American Journal of Physics 84(5), pp. 327–334, 2016.

Harvey R. Brown and James Read, "The Dynamical Approach to Spacetime", in E. Knox and A. Wilson (eds.), *The Routledge Companion to Philosophy of Physics*, Oxford: Routledge, 2020.

Jeremy Butterfield, "Reconsidering Relativistic Causality", International Studies in the Philosophy of Science 21(3), pp. 295–328, 2007.

Sean Carroll and Eugene Lim, "Lorentz-Violating Vector Fields Slow the Universe Down", Physical Review D 70, 123525, 2004.

A. Coley, "Schiff's Conjecture on Gravitation", Physical Review Letters 49(12), pp. 853–855, 1982.

Erik Curiel, "A Primer on Energy Conditions", in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, pp. 43–104, 2017.

Erik Curiel, "On Geometric Objects, the Non-Existence of a Gravitational Stress-Energy Tensor, and the Uniqueness of the Einstein Field Equation", Studies in History and Philosophy of Modern Physics 66, pp. 90–102, 2019.

Neil Dewar, "Maxwell Gravitation", Philosophy of Science 85, pp. 249–270, 2018.

Neil Dewar and James Owen Weatherall, "On Gravitational Energy in Newtonian Theories", Foundations of Physics 48(5), pp. 558–578, 2018.

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

Jsürgen Ehlers and Robert Geroch, "Equation of Motion of Small Bodies in Relativity", Annals of Physics 309, pp. 232–236, 2004.

Jürgen Ehlers, Felix A. E. Pirani and Alfred Schild, "The Geometry of Free Fall and Light Propagation", in L. O'Reifeartaigh (ed.), *General Relativity: Papers in Honour of J. L. Synge*, Oxford: Clarendon Press, pp. 63–84, 1972.

Michael Friedman, *Foundations of Space-Time Theories*, Princeton: Princeton University Press, 1983.

Robert Geroch and Pong Soo Jang, "Motion of a Body in General Relativity", Journal of Mathematical Physics 16(1), pp. 65–67, 1975.

Robert Geroch and James Owen Weatherall, "The Motion of Small Bodies in Space-Time", Communications in Mathematical Physics 364, pp. 607–634, 2018.

Carl Hoefer, "Energy Conservation in GTR", Studies in History and Philosophy of Modern Physics 31, pp. 187–199, 2000.

Nick Huggett, "Essay Review: *Physical Relativity* and *Understanding Space-Time*", Philosophy of Science 76, pp. 404–422, 2009.

Ted Jacobson and David Mattingly, "Gravity with a Dynamical Preferred Frame", Physical Review D 64, 024028, 2001.

Michel Janssen,"COI Stories: Explanation and Evidence in the History of Science", Perspectives on Science 10, pp. 457–522, 2002.

Michel Janssen, "Drawing the Line Between Kinematics and Dynamics in Special Relativity", Studies in History and Philosophy of Modern Physics 40, pp. 26–52, 2009.

Eleanor Knox, "Effective Spacetime Geometry", Studies in History and Philosophy of Modern Physics 44, pp. 346–356, 2013.

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

Eleanor Knox, "Physical Relativity from a Functionalist Perspective", Studies in History and Philosophy of Modern Physics 67, pp. 118–124, 2019.

Vincent Lam, "Gravitational and Nongravitational Energy: The Need for Background Structures", Philosophy of Science 78, pp. 1012–1024, 2011.

Dennis Lehmkuhl, "Mass-Energy-Momentum: Only there Because of Spacetime?", British Journal for the Philosophy of Science 62, pp. 453–488, 2011.

Alan P. Lightman and David L. Lee, "Restricted Proof that the Weak Equivalence Principle Implies the Einstein Equivalence Principle", Physical Review D 8(2), pp. 364–376, 1973.

David Malament, "A Remark About the "Geodesic Principle" in General Relativity", in M. Frappier, D. H. Brown and R. DiSalle (eds.), *Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos*, New York: Springer, pp. 245–252, 2012.

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

Tim Maudlin, *Philosophy of Physics Volume I: Space and Time*, Princeton: Princeton University Press, 2012.

Tushar Menon, Niels Linnemann and James Read, "Clocks and Chronogeometry: Rotating Spacetimes and the Relativistic Null Hypothesis", British Journal for the Philosophy of Science, 2018 (forthcoming).

Hermann Minkowski, "Raum und Zeit", Physikalische Zeitschrift 10, pp. 104–111, 1909.

Charles Misner, Kip Thorne, and John Wheeler, *Gravitation*, San Francisco: Freeman & Co., 1973.

Wayne C. Myrvold, "How Could Relativity be Anything Other Than Physical?" Studies in History and Philosophy of Modern Physics 67, pp. 137–143, 2019.

Wei-Tou Ni, "Equivalence Principles and Electromagnetism", Physical Review Letters 38(7), pp. 301–304, 1977.

John D. Norton, "Why Constructive Relativity Fails", British Journal for the Philosophy of Science 59, pp. 821–834, 2008.

J. Brian Pitts, "Absolute Objects and Counterexamples: Jones-Geroch Dust, Torretti Constant Curvature, Tetrad-Spinor, and Scalar Density", Studies in History and Philosophy of Modern Physics 37(2), pp. 347–371, 2006.

J. Brian Pitts, "Gauge-Invariant Localization of Infinitely Many Gravitational Energies from All Possible Auxiliary Structures", General Relativity and Gravitation 42, 2010.

J. Brian Pitts, "Space-time Constructivism *vs*. Modal Provincialism: Or, How Special Relativistic Theories Needn't Show Minkowski Chronogeometry", Studies in History and Philosophy of Modern Physics 67, pp. 191–198, 2019.

Oliver Pooley, "Substantivalist and Relationist Approaches to Spacetime", in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, 2013.

Oliver Pooley, "Background Independence, Diffeomorphism Invariance, and the Meaning of Coordinates", in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, pp. 105–144, 2017.

James Read, "Functional Gravitational Energy", British Journal for the Philosophy of Science 71, pp. 205–232, 2020.

James Read, "On Miracles and Spacetime", Studies in History and Philosophy of Modern Physics 65, pp. 103–111, 2019.

James Read, Harvey R. Brown and Dennis Lehmkuhl, "Two Miracles of General Relativity", Studies in History and Philosophy of Modern Physics 64, pp. 14–25, 2018.

Simon Saunders, "Rethinking Newton's *Principia*", Philosophy of Science 80, pp. 22–48, 2013.

L. I. Schiff, "On Experimental Tests of the General Theory of Relativity", American Journal of Physics 28, pp. 340–343, 1960.

Howard Stein, "Some Philosophical Prehistory of General Relativity", in J. Earman, C. Glymour and J. Stachel (eds.), *Foundations of Space-Time Theories*, Minnesota Studies in the Philosophy of Science, volume 8, Minneapolis: University of Minnesota Press, pp. 3–49, 1977.

Syman Stevens, "The Dynamical Approach as Practical Geometry", Philosophy of Science 82, pp. 1152–1162, 2015.

Syman Stevens, "Regularity Relationalism and the Constructivist Project", British Journal for the Philosophy of Science 71, pp. 205–232, 2020.

Nicholas J. Teh, “Recovering Recovery: On the Relationship Between Gauge Symmetry and Trautman Recovery”, Philosophy of Science 85, pp. 201–224, 2018.

Kip S. Thorne, David L. Lee, and Alan P. Lightman, “Foundations For a Theory of Gravitation Theories”, Physical Review D 7, pp. 3563–3578, 1973.

David Wallace, “Fundamental and Emergent Geometry in Newtonian Physics”, British Journal for the Philosophy of Science 71, pp. 1–32, 2020.

James Owen Weatherall, “On (Some) Explanations in Physics”, Philosophy of Science 78(3), pp. 421–447, 2011.

James Owen Weatherall, “On the Status of the Geodesic Principle in Newtonian and Relativistic Physics”, Studies in the History and Philosophy of Modern Physics 42(4), pp. 276–281, 2011.

James Owen Weatherall, “The Motion of a Body in Newtonian Theories”, Journal of Mathematical Physics 52(3), 032502, 2011.

James Owen Weatherall, “A Brief Remark on Energy Conditions and the Geroch-Jang Theorem”, Foundations of Physics 42(2), pp. 209–214, 2012.

James Owen Weatherall, “Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?”, Erkenntnis 81(5), pp. 1073–1091, 2016.

James Owen Weatherall, “Maxwell-Huygens, Newton-Cartan, and Saunders-Knox Space-Times”, Philosophy of Science 83, pp. 82–92, 2016.

James Owen Weatherall, “Understanding Gauge”, Philosophy of Science 85(5), pp. 1039–1049, 2016.

James Owen Weatherall, “A Brief Comment on Maxwell(/Newton)[-Huygens] Spacetime”, Studies in History and Philosophy of Modern Physics 63, pp. 34–38, 2018.

James Owen Weatherall, “Conservation, Inertia, and Spacetime Geometry”, Studies in History and Philosophy of Modern Physics 66, pp. 144–159, 2019.

James Owen Weatherall, “Inertial Motion, Explanation, and the Foundations of Classical Spacetime Theories”, in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, pp. 13–42, 2017.

James Owen Weatherall, “Categories and the Foundations of Classical Field Theories”, in E. Landry (ed.), *Categories for the Working Philosopher*, pp. 329–348, Oxford: Oxford University Press, 2018.

Hermann Weyl, “Zur Infinitesimalgeometrie: Einordnung der Projektiven und der Konformen Auffasung”, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, Göttingen, pp. 99–112, 1921.

Hermann Weyl, *Mathematische Analyse des Raumproblems*, lecture 3, Berlin, 1923.

Geometry and Motion in General Relativity

James Owen Weatherall

Abstract A classic problem in general relativity, long studied by both physicists and philosophers of physics, concerns whether the geodesic principle may be derived from other principles of the theory, or must be posited independently. In a recent paper [Geroch & Weatherall, "The Motion of Small Bodies in Space-Time," *Comm. Math. Phys.* (forthcoming)], Bob Geroch and I have introduced a new approach to this problem, based on a notion we call "tracking." In the present paper, I situate the main results of that paper with respect to two other, related approaches, and then make some preliminary remarks on the interpretational significance of the new approach. My main suggestion is that "tracking" provides the resources for eliminating "point particles"—a problematic notion in general relativity—from the geodesic principle altogether.

1 Introduction

There is a deep link in general relativity between, on the one hand, the geometry of spacetime and, on the other hand, the motion of small bodies. Spacetime in the theory is represented by a smooth manifold M endowed with a smooth metric g_{ab}; this metric (and its associated Levi-Civita derivative operator, ∇) determine a class of *timelike geodesics*, which are the curves of "locally extremal" length. These curves have special physical significance in the theory: they are the possible trajectories of free massive test point particles. Thus we have an identification between a class of geometrically privileged curves and a class of physically privileged trajectories in general relativity.

This link is sometimes called the *geodesic principle*; it is analogous to Newton's first law of motion. Because of its centrality to the interpretation of spacetime geometry in general relativity, this principle has received a great deal of attention

J. O. Weatherall (✉)
Department of Logic and Philosophy of Science, University of California, Irvine, CA, USA
e-mail: weatherj@uci.edu

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_10

from both physicists and philosophers of physics, going back at least to Einstein and Grommer (1927).[1] One issue of particular significance concerns whether the geodesic principle is an independent postulate, or if, instead, it should be understood as a consequence of other principles of the theory. This question is known as the "problem of motion" in general relativity. In fact, it is widely recognized that the geodesic principle *is*, in some sense, a consequence of the rest of the theory. But articulating this sense in a satisfactory way turns out to be remarkably subtle. Over the past century, dozens of different attempts have been made to capture, in a mathematically precise and physically perspicuous way, the sense in which the geodesic principle is a theorem of general relativity.

In a recent paper, Bob Geroch and I have introduced a new approach to the problem of motion (Geroch and Weatherall, 2018). The main theorem of that paper, Theorem 4 below, captures a sense in which generic small bodies in general relativity follow timelike geodesics (and light rays follow null geodesics). This theorem has a number of virtues over other approaches, at least some of which are salient to recent discussions in the philosophy of physics literature concerning the relationship between spacetime geometry and the dynamics of matter. My goal here is to present the main results of that earlier paper in a way that emphasizes some of these relative virtues, and then make some preliminary remarks on the interpretational significance of the results. The main suggestion—which is only implicit in the earlier paper—will be that the methods used in stating and proving this theorem provide the resources for eliminating "point particles"—a problematic notion in general relativity—from the geodesic principle altogether. Instead, I will argue that one can capture the substance of the link between geometry and motion directly as an assertion about the solutions to the field equations governing realistic, extended matter.

The remainder of the paper will proceed as follows. In the next section, I will describe two well-known approaches to the problem of motion, and discuss some of their shortcomings. In the following section I will present the main results from Geroch and Weatherall (2018) and explain how they combine the virtues of these two other approaches while avoiding their problems. I will conclude by discussing how one might re-think the geodesic principle in light of these results. I emphasize that I will not attempt to reproduce the discussion in Geroch and Weatherall (2018), and I direct the reader there for many of the mathematical details and for proofs of propositions. Rather, my goal is to give a different, complementary presentation of (some of) that material, with an emphasis on its motivation, what makes it distinctive, and some of the reasons why it might be of interest to philosophers.

[1]For a recent review of the physics literature on this subject, see Poisson et al. (2011); for other recent work, see Asada et al. (2011), Gralla and Wald (2011), and the contributions to Puetzfeld et al. (2015). For the recent philosophical literature, which generally stems from a discussion of the geodesic principle by Brown (2005), see Malament (2012a), Tamir (2012), Sus (2014), Samaroo (2018), Lehmkuhl (2017b,a), and Weatherall (2011b, 2017, 2019).

2 Two Approaches and Their Discontents

In what follows, fix a relativistic spacetime, (M, g_{ab}).[2] The geodesic principle states that free massive test point particles traverse timelike geodesics of spacetime. In this section, I present two widely discussed approaches to capturing the geodesic principle as a theorem of general relativity, and describe reasons why one might be dissatisfied with each of them. I do not mean to claim that these are the *only* two approaches in the literature—to the contrary, there are many approaches out there.[3] But these two are distinguished by the fact that they yield precise mathematical theorems that are strong and simple, and do not rely on physical arguments that call into question the generality of the results.[4]

To begin, however, let me comment on (part of) why formulating such theorems requires care. Basically, the difficulty comes down to the fact that "free massive test point particles" are not particularly natural objects in general relativity. For reasons I discuss below, one usually considers extended matter, represented by smooth fields of various kinds. In principle, one would like to say that a point particle is an idealization of a "small body," and so one would like to associate small extended bodies with curves that they "traverse." In special relativity, as in Newtonian mechanics, there is no difficulty in doing so: one can identify, with any extended body, a unique "center of mass" trajectory, reflecting the "average" motion of the body, and then argue that that trajectory must be a timelike geodesic.[5]

But in curved spacetime, analogous constructions are apparently not possible. In that context, although there are of course many curves that lie within the worldtube of any given (extended) body, it is not clear that any of them captures the overall

[2] Although this is meant to be a relatively gentle introduction, I take for granted the basic mathematics of general relativity; for relevant background, see, for instance, Wald (1984) or Malament (2012b), both of whom use essentially the same notation as I do.

[3] Once again, see the references in note 1. One approach in particular that has been widely influential, but which I do not discuss at all, is the method of matched asymptotic expansion, as developed, for instance, by D'Eath (1975), Thorne and Hartle (1985), Mino et al. (1997), and Gralla and Wald (2011).

[4] There is a certain trade-off between, on the one hand, strength and simplicity and, on the other hand, information relevant in special cases, such as possible deviations from geodesic motion that might arise from finite body effects "on the way to the limit." Compare, for instance, Thorne and Hartle (1985) or Gralla and Wald (2011), who describe, in the presence of additional (strong) assumptions, higher order "corrections" to geodesic motion for finite bodies, with the results to be described here, which might be understood to characterize (without these strong assumptions) the universal limiting, or order zero, behavior of small bodies. My perspective is that for foundational purposes, the more general and precise results are of greater value, though this is not necessarily true for other purposes, such as studying binary black holes. On the other hand, see footnotes 10 and 22 for ways in which the perspective taken here may bear fruitfully on widely accepted results from other approaches.

[5] Geroch and Jang (1975) give a compact treatment of the situation in special relativity. For further discussion of the situation regarding theorems of the present sort in Newtonian gravitation, see Weatherall (2011a,b, 2017). I will not discuss these results further in the present paper.

motion of the body—and in general, there need not be *any* geodesic lying within the (convex hull) of the worldtube of a body, even in the absence of external forces. This suggests that, whatever else is the case, the geodesic principle should only hold in the limit as the radius of a body goes to zero; for extended matter in curved spacetime, it is hard to identify even a candidate assertion that captures the idea that bodies move on geodesics.

2.1 Distributions

One way to overcome these challenges is to give up on representing bodies with smooth fields, and instead to consider point particles represented as *distributions*—basically, generalized functions, such as δ functions and their derivatives—that are supported only on curves. Roughly, a distribution is a map from a space of *test fields*, which are usually smooth fields of compact support, to the real numbers that is continuous in a suitable sense.[6] Although they are not smooth functions, one can generally manipulate them as if they were, for instance, by taking their derivatives.[7] As first observed by Mathisson (1931), and developed by Souriau (1974), Sternberg and Guillemin (1984), and others, this approach leads to a very short argument for geodesic motion.

The argument goes as follows. Suppose one is given a symmetric distribution $\mathbf{T}^{ab}$ supported on a timelike curve γ in (M, g_{ab}). We might take this distribution to represent the energy-momentum of a small body—or better, a point particle, since it has no spatial extension. Now suppose this distribution is order zero and divergence-free, where by order zero we mean the action of $\mathbf{T}^{ab}$ may be extended from smooth fields of compact support to merely continuous fields (which means, roughly, that the value $\mathbf{T}^{ab}$ yields when acting on a test field α_{ab} depends only on the value of α_{ab} at each point, and not its derivatives), and by divergence-free, we mean that $\nabla_a \mathbf{T}^{ab} = \mathbf{0}$.[8] It follows, by a short calculation, that $\mathbf{T}^{ab} = m\boldsymbol{\delta}_\gamma u^a u^b$, where m is a number, $\boldsymbol{\delta}_\gamma$ is the delta distribution supported on γ, and u^a is the unit tangent to γ. It also follows that γ is a (timelike) geodesic.

This approach has some obvious advantages. The argument just given is mathematically very simple. It is also easy to generalize to forces. For instance, still assuming $\mathbf{T}^{ab}$ order zero, a body with timelike worldline γ, subject to an arbitrary

[6]More precisely, we take test fields to be *densities* of weight 1; see Geroch and Weatherall (2018, Appendix A).

[7]We take derivatives by analogy with integration by parts. Fix a manifold M, a derivative operator ∇ on M, and a distribution $\mathbf{X}$ on M. Then $\nabla_a \mathbf{X}$ is that distribution whose action on a smooth test field α^a is given by $\nabla_a \mathbf{X}\{\alpha^a\} = -\mathbf{X}\{\nabla_a \alpha^a\}$. For background on distributions, including tensor distributions, see Geroch and Weatherall (2018, Appendix A), Grosser et al. (2001), or Steinbauer and Vickers (2006). The details of the theory of distributions do not particularly matter for the arguments that follow.

[8]That is, $\mathbf{T}^{ab}$ vanishes on all test fields of the form $\nabla_a \alpha_b$.

force $\mathbf{f}^a = \nabla_b \mathbf{T}^{ab}$, can be described by an energy-momentum $\mathbf{T}^{ab} = \boldsymbol{\mu} u^a u^b$ satisfying

$$\boldsymbol{\mu} u^n \nabla_n u^a = q^a{}_b \mathbf{f}^b$$
$$\nabla_b(\boldsymbol{\mu} u^b) = -\mathbf{f}^b u_b,$$

where $\boldsymbol{\mu}$ is an order zero distribution supported on γ. The first of these equations asserts precisely that $\mathbf{F} = m\mathbf{a}$; and the second is a "continuity" equation describing the possibility of transfer of mass between different bodies.

Likewise, fix a background electromagnetic field F_{ab}. Represent a charged body by an energy-momentum distribution $\mathbf{T}^{ab}$ supported on a timelike curve γ and an (order zero) charge-current density $\mathbf{J}^a$, also supported on γ. Assume $\nabla_a \mathbf{J}^a = 0$ and $\mathbf{f}^a = F^a{}_b \mathbf{J}^b$. Then $\mathbf{J}^a = e\boldsymbol{\delta}_\gamma u^a$, $\mathbf{T}^{ab} = m\boldsymbol{\delta}_\gamma u^a u^b$, and γ is a e/m Lorentz force curve, that is, $u^n \nabla_n u^a = e/m F^a{}_b u^b$. Moreover, one can solve for the general case, where $\mathbf{J}^a$ is order one (the highest order compatible with $\mathbf{T}^{ab}$ being order zero); one finds contributions to the motion arising from electric and magnetic dipoles.

So distributions do not merely capture the idea of "free" motion; they also allow us to derive general claims about particle motion, including the Lorentz force law. But despite this simplicity and power, the situation concerning distributions is not entirely satisfactory.

One concern is immediate. We assumed, from the beginning, that the distribution $\mathbf{T}^{ab}$ representing the energy-momentum of a point particle is order zero. Without this assumption, none of the arguments above go through, and indeed, one can find divergence-free distributions on any curve at all.[9] But why assume this?

In fact, the restriction to order zero distributions can be justified by the following argument. Let us say that a smooth test field t_{ab} satisfies the *dual energy condition* at a point p if t_{ab} can be written as a sum of symmetrized outer products of co-oriented causal covectors. The fields satisfying this condition at a point are precisely the ones that are "dual" to tensors T^{ab} satisfying the (standard) dominant energy condition, which states that given any pair of co-oriented causal vectors η^a and ξ^a, $T^{ab}\xi_a\eta_b \geq 0$. We then say that a symmetric distribution $\mathbf{T}^{ab}$ satisfies the *dominant energy condition* if, for every test field t_{ab} satisfying the dual energy condition, $\mathbf{T}^{ab}\{t_{ab}\} \geq 0$. Note that this is a straightforward extension of the dominant energy condition from tensors at a point (and smooth tensor fields) to distributions.

We then get the following result.

Proposition 1 *Let $\mathbf{T}^{ab}$ be a symmetric distribution satisfying the dominant energy condition. Then $\mathbf{T}^{ab}$ is order zero.*

[9] This is the distributional analog to the result proved in Malament (2012a).

Thus, insofar as one expects matter to satisfy the dominant energy condition, it should be represented by distributions that are order zero.[10]

But other concerns are less easily dealt with. In particular, although distributions seem like a natural way of representing "point particles" in general relativity, it is difficult to see how they are related to "realistic" matter. As I noted above, matter in general relativity is usually represented by smooth fields. These fields are generally solutions to certain systems of partial differential equations, such as Maxwell's equations or the Klein–Gordon equation; each such solution is associated with some (smooth) energy-momentum tensor, via standard formulae. For standard examples (Maxwell, Klein–Gordon, etc.), energy-momentum tensors are quadratic in field values and/or their derivatives.

But this commonplace observation is a big problem for the distributional approach. If we consider only smooth solutions to these equations, then the associated energy-momentum tensors will also be smooth, i.e., they will not be distributions supported on a curve. So the distributional energy-momenta considered above cannot arise in this way. One might think that this means we should consider distributional solutions to the matter field equations, in which case one could perhaps find solutions that are supported on a curve. But even if one had such a solution, one could not generally associate an energy-momentum tensor with it. The reason is that multiplication of distributions is not well-defined.[11] And so it is not clear how the distributional energy-momenta we have been considering are supposed to arise, or what kind of matter they represent.

A related difficulty arises when we try to understand distributional energy-momenta as sources in Einstein's equation. In fact, a well-known result due to Geroch and Traschen (1987) establishes that there are no metrics satisfying certain weak conditions compatible with distributional sources supported on a curve. Thus it is difficult to evaluate, for instance, backreaction arising from a distributional $\mathbf{T}^{ab}$.

And so it seems that distributional energy-momenta cannot arise from realistic matter, and they cannot act as sources in Einstein's equation. So in what sense do distributional $\mathbf{T}^{ab}$ represent anything physical? And what bearing do the simple results described above have on the motion of actual bodies?

[10] Although it is a side issue for present purposes, observe that this result points to a problem with certain approaches to treating the motion of rotating particles that represent "spin" by higher order distributions supported on a curve (Papapetrou, 1951; Souriau, 1974): such particles are incompatible with the energy condition. There is good physical reason for this. For ever smaller bodies to have large angular momentum (per unit mass), their rotational velocity must increase without bound—leading to superluminal velocities, which are incompatible with the energy condition.

[11] There are extensions to the theory of distributions—namely, the theory of Colombeau algebras (Colombeau, 2000)—that permit one to multiply distributions. But these have some undesirable properties, including that multiplication is not uniquely defined for distributions, and it does not reduce to pointwise multiplication for all (continuous) functions, conceived as distributions.

2.2 Curve-First

A second approach to the problem of motion, which has been widely discussed in the philosophical literature, was developed by Geroch and Jang (1975) and Ehlers and Geroch (2004). On this approach, one begins with a curve γ and considers smooth fields T^{ab}, satisfying the dominant energy condition, supported in small neighborhoods of the curve. These fields represent the energy-momenta of small bodies propagating "near" the curve γ. One then proves the following theorem.

Theorem 2 (Geroch–Jang) *Let γ be a smooth, timelike curve in a spacetime (M, g_{ab}). Suppose that, in any neighborhood O of γ, there exists a smooth, symmetric, divergence-free, and non-vanishing tensor field T^{ab} satisfying the dominant energy condition whose support lies in O. Then γ is a geodesic.*[12]

The interpretation of this result is perhaps not quite as straightforward as the distributional result. The idea is that the only curves along which arbitrarily small massive bodies (represented by spatially localized T^{ab} fields, satisfying the dominant energy condition) may propagate in the absence of any external forces (captured here by the requirement that the fields be divergence-free) are (timelike) geodesics. Thus we get a sense in which free massive point particles must follow timelike geodesics.[13]

Like the distributional approach, the curve-first approach is also simple. And since it refers only to smooth T^{ab} fields, its physical interpretation is more transparent. Moreover, smooth T^{ab} fields may be sources in Einstein's equation, and so this method may be adapted to consider backreaction. Indeed, there is a strengthening of the Geroch–Jang theorem that captures precisely this:

Theorem 3 (Ehlers–Geroch) *Let γ be a smooth, timelike curve in a spacetime (M, g_{ab}). Suppose that, for any (closed) neighborhood O of γ, and any $C^1[O]$ neighborhood $\hat{O}$ of g_{ab}, there exists a Lorentzian metric $\hat{g} \in \hat{O}$ whose Einstein tensor is non-vanishing, which satisfies the dominant energy condition (relative to g_{ab}), and whose support lies in O. Then γ is a geodesic.*

The interpretation of this result is that even if we consider small bodies that "perturb" the spacetime metric g_{ab}, at least to first order, in the limit as those bodies become small (in mass and spatial extent), they must follow timelike geodesics of g_{ab}.

But again, the situation is not totally satisfactory. One issue is that curve-first results work well for *free* bodies, but it is difficult to see how to generalize them to include forces—including, for instance, the Lorentz force law, which one might

[12]Observe that we assume from the start that the curve is timelike; if one wants to *conclude* that the curve must be timelike, a stronger energy condition is required (Weatherall, 2012).

[13]For further discussion of the interpretation of this theorem, see Weatherall (2011b, 2017); I do not wish to belabor here points I already make elsewhere.

have guessed would have a similar status as the geodesic principle.[14] Recall that in the distributional case, these sorts of generalizations were almost immediate, because the energy condition placed a strong constraint on possible forces and also on, for instance, charge-current densities. But on the curve-first approach, the energy condition does not seem to place analogous constraints on smooth T^{ab} fields. It seems some further conditions are needed to recover the equations characterizing forces.

Another concern is that, although curve-first results consider smooth fields, there is still a problem concerning "realistic" matter, in the form of solutions to some hyperbolic system. The issue now has to do with the way in which the limit is taken. In particular, the Geroch–Jang and Ehlers–Geroch theorems assume matter fields can vanish outside of arbitrary neighborhood of a timelike curve. But for hyperbolic systems, this is not generally possible: solutions to the Maxwell or Klein–Gordon equations, for instance, tend to spread over time, and there are, in general, no solutions that are supported arbitrarily closely to a curve for all time. This leads to the following embarrassing situation: the geodesic principle theorems do not establish that Maxwell fields follow null geodesics, even in an appropriate high-frequency (optical) limit![15]

Thus we find that curve-first results, like distributional results, are of limited physical applicability. In particular, it is not clear how to think of solutions of the field equations that govern real matter in general relativity as somehow realizing the conditions assumed in the limiting procedure for these results.

3 The Miracle of Tracking

We saw in the last section that both the distributional and curve-first approaches have some attractive features—but that neither is fully satisfactory. In this section, I describe a novel approach to the problem that combines the distributional and curve-first approaches, and does so in a way that allows us to extend both while also clarifying the physical significance of both constructions.[16] The results here are

[14] Gralla et al. (2009) extend a version of a curve-first approach to treat the Lorentz force law, and also derive leading order "self-force" corrections to it. But the relationship between their arguments and the sort of result envisaged here is the same as the relationship between the Gralla and Wald (2011) results and, say, the Geroch and Jang (1975) theorem, which is that they require much stronger assumptions. (Recall footnote 4.)

[15] For instance, in his classic textbook Wald (1984) describes the Geroch–Jang theorem as capturing the sense in which small bodies follow timelike geodesics, but then does not invoke this result to establish that light rays traverse null geodesics—appealing, instead, to a completely different construction.

[16] There is a sense in which Gralla and Wald (2011) and Gralla et al. (2009) also combine features of both approaches, though their approach is considerably different. Recall, again, footnote 4.

from Geroch and Weatherall (2018); proofs of all propositions, as well as further discussion emphasizing different issues, can be found there.

3.1 Definition of Tracking

The key concept in this approach is that of *tracking*, which we introduce now. To begin, fix, once again, a relativistic spacetime (M, g_{ab}). Let us suppose that we are given, on this spacetime, a collection $\mathcal{C}$ of smooth, symmetric fields T^{ab}, each satisfying the dominant energy condition. Although these fields are smooth, each of them is naturally associated with a distribution, the action of which on test fields x_{ab} is defined by $\mathbf{T}^{ab}\{x_{ab}\} = \int_M T^{ab} x_{ab}$. We will say that this collection *tracks* a timelike curve γ if, for every smooth test field x_{ab} satisfying the dual energy condition in a neighborhood of γ and generic at some point of γ,[17] there is a field T^{ab} in $\mathcal{C}$ such that $\mathbf{T}^{ab}\{x_{ab}\} > 0$.[18]

The rest of this section concerns facts about, and applications of, tracking. Since this concept is the main idea in what follows, its interpretation demands special attention. First, observe that because each field T^{ab} in the collection $\mathcal{C}$ satisfies the dominant energy condition, when you contract it, at a point, with a test field that satisfies the dual energy condition there, the result is non-negative (Observe that this makes sense, since the fields in $\mathcal{C}$ are ordinary smooth tensor fields; they determine distributions, but we can also consider their action on vectors and covectors at a point.). This means that when a field in $\mathcal{C}$ acts, as a distribution, on a test field that satisfies the dual energy condition everywhere, then the result is necessarily non-negative (though it may vanish). (This, recall, is just what it means to say that a distribution satisfies the dominant energy condition.) We may thus think of any given test field x_{ab}, satisfying the dual energy condition everywhere, as giving a standard of "magnitude" for T^{ab} in the region where x_{ab} is supported, with different test fields giving different standards.

Of course, none of this holds if one acts on test fields that satisfy the dual energy condition only at some points—in that case, fields in $\mathcal{C}$ may or may not yield a non-negative result. Given a test field x_{ab}, however, satisfying the dual energy condition in a region O (and non-vanishing there), one can always construct a field T^{ab}, satisfying the dominant energy condition, whose action, as a distribution, on x_{ab} is positive, by ensuring that T^{ab} is sufficiently "large" in O and "small" in $M - O$ (by the standard of "large" and "small" given by x_{ab}). That is, one can choose T^{ab} so that the part of the integral taken over O dominates, i.e., so that

[17] By "generic" at a point p, I mean that x_{ab} lies in the interior of the cone of tensors satisfying the dual energy condition at a p: that is, for any non-vanishing tensor T^{ab} satisfying the dominant energy condition at p, $T^{ab}x_{ab} > 0$.

[18] Observe the notational convention adopted here: previously we had used boldface for distributions; now we are using bold symbols to refer to the distributions associated with (determined by) smooth fields represented by the same, non-bolded, symbol.

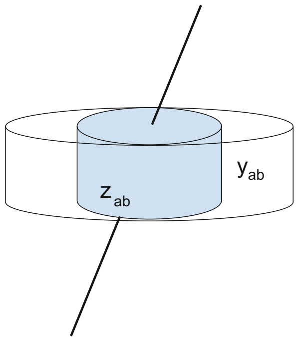


Fig. 1 Here we depict the basic construction underlying the notion of "tracking." Consider two test fields, z_{ab} and y_{ab}, both satisfying the dual energy condition, but where z_{ab} is supported near a curve γ and y_{ab} is supported away from γ. Then $x_{ab} = z_{ab} - y_{ab}$ satisfies the dominant energy condition near γ. If T^{ab} satisfying the dominant energy condition satisfies $T^{ab}\{x_{ab}\} > 0$, then there is "more" T^{ab} in the region where z_{ab} is supported than the region where y_{ab} is supported

$$\int_O T^{ab}x_{ab} \geq \left| \int_{M-O} T^{ab}x_{ab} \right|.$$

Indeed, this interpretation is particularly clear in cases where a test field x_{ab} may be decomposed as the difference of two test field y_{ab} and z_{ab}, i.e., $x_{ab} = z_{ab} - y_{ab}$, both satisfying the dual energy condition everywhere. In that case, we have

$$\int_M T^{ab}x_{ab} = \int_M T^{ab}z_{ab} - \int_M T^{ab}y_{ab},$$

which, since both integrals on the right-hand side are always non-negative, yields a positive number if and only if there is "more" T^{ab} in the region of support of z_{ab} (by the standard given by z_{ab}) than there is in the region of support of y_{ab} (again, by the standard given by y_{ab}) (See Fig. 1.).

With these remarks in hand, we return to the definition of tracking. There we require that, for *any* test field x_{ab}, satisfying the dual energy condition in a neighborhood of γ, there is an element of $\mathcal{C}$ whose action on x_{ab} is (strictly) positive. This captures the idea that, by any standard one likes—or at least, any standard captured by test fields—for measuring "amount of T^{ab} near γ" and "amount of T^{ab} away from γ," T^{ab} may be chosen from $\mathcal{C}$ so that there is more T^{ab} near γ than away from γ—or in other words, there exist fields T^{ab} in $\mathcal{C}$ that are as concentrated as one wishes near γ. Note that we consider only those "standards of measurement" given by test fields, which always have compact support. This means that we are considering T^{ab} fields that are arbitrarily concentrated near γ for arbitrarily long, but finite, duration. It also means that there could be arbitrarily large amounts of T^{ab} far from γ, as long as it does not fall within the support of x_{ab}.

3.2 Consequences of Tracking

As we have just seen, tracking gives us a sense in which a collection of fields includes elements that follow a curve γ as closely as one likes, for as long as one likes. It then follows that a collection $\mathcal{C}$, satisfying various properties, can track only certain curves. In particular, we get the following result.[19]

Theorem 4 *Let (M, g_{ab}) be a spacetime, γ a timelike curve therein, and $\mathcal{C}$ a collection of symmetric fields T^{ab}, each satisfying the dominant energy condition, that tracks γ. Suppose each of these fields is conserved. Then there exists a sequence of fields $T_1^{ab}, T_2^{ab}, \ldots$, each a positive multiple of some element of $\mathcal{C}$, that converges, in the sense of distributions, to $\boldsymbol{\delta}_\gamma u^a u^b$.*

Corollary 5 *The curve γ is a geodesic.*

With a small modification, Theorem 4 and Corollary 5 hold for null curves as well.[20] We also have the following converse:[21]

Theorem 6 *Let (M, g_{ab}) be a spacetime, γ a curve therein, and $\mathcal{C}$ a collection of symmetric fields T^{ab}, each satisfying the dominant energy condition that tracks γ. Then γ is timelike or null.*

These results imply that the only curves that collections of smooth, symmetric, rank 2 fields, all divergence-free and satisfying the dominant energy condition, can track are timelike or null geodesics. But they also say more than this: they assert that *any* family of bodies, satisfying the energy condition and the conservation condition, that follows a curve arbitrarily tightly contains a sequence converging, up to rescaling, to the (distributional) energy-momentum representing a point particle. In other words, *every* sequence of smooth, symmetric, divergence-free fields, satisfying the dominant energy condition, whose support approaches a timelike or null curve γ, converges, up to rescaling, to a multiple of the δ distribution on γ. This captures the sense in which the distribution $\boldsymbol{\delta}_\gamma u^a u^b$ represents the energy-momentum of

[19]These results rephrase Theorem 3 and the subsequent discussion of Geroch and Weatherall (2018).

[20]The small modification involves the definition of a δ distribution supported on a null curve, which requires a choice of parameterization (since null curves cannot be parameterized by arc-length). It does not affect the conclusion that γ is a geodesic.

[21]Observe that this converse may be understood to capture a sense in which superluminal propagation is impossible, at least in a point-particle limit. One might take this result to be in tension with the arguments of Geroch (2011) and Weatherall (2014). But in fact, the tension is only apparent: this result assumes the dominant energy condition, while the discussions in those other papers do not (see also Earman (2014) for a discussion of the relation between the dominant energy condition and the notion of "superluminal propagation" discussed there). That said, the present result, in connection with those earlier papers, raises an interesting question. Can one generalize the notion of tracking considered here to hyperbolic systems whose solutions do not satisfy the dominant energy condition, and if so, do solutions always track their characteristics? I am grateful to an anonymous referee for raising the possible tension.

realistic (extended) matter: it is the essentially unique accumulation point for energy-momentum tensors of small bodies. The key insight is that bodies that can be made arbitrarily small in size, in the sense captured by tracking, necessarily approach delta functions on a curve.[22]

So Theorem 4 makes direct contact with the distributional results described in Sect. 2, and it clarifies the physical significance of order zero, divergence-free distributions $\mathbf{T}^{ab}$ supported on a curve. But this theorem *also* captures, and indeed strengthens, the Geroch–Jang result. To see this, observe the following:

Proposition 7 *If $\mathcal{C}$ contains, for every neighborhood O of a curve γ, a smooth, symmetric, non-vanishing, divergence-free field T^{ab} that satisfies the dominant energy condition and vanishes outside of O, then $\mathcal{C}$ tracks γ.*

In other words, the antecedent of the Geroch–Jang theorem implicitly defines a collection $\mathcal{C}$ of fields T^{ab}, each satisfying the dominant energy condition and each divergence-free: these are the fields that are supported (only) in arbitrarily small neighborhoods of a curve γ. What this proposition asserts is that this collection tracks γ; it follows, then, from Theorem 4 that not only is γ a geodesic, but that the collection $\mathcal{C}$ contains a sequence that, up to rescaling, converges to a δ distribution on γ. Moreover, we see how the collection $\mathcal{C}$ defined by the Geroch–Jang theorem is more restrictive than necessary to get this result—and thus we see the sense in which Theorem 4 is a strict strengthening of the Geroch–Jang theorem. One can recover the Ehlers–Geroch theorem in a similar manner.

3.3 Applications of These Results

The results just described allow us to extend the curve-first approach in two important ways. First, by connecting curve-first and distributional results, Theorem 4 provides an important hint on how to extend the curve-first approach to forced motion. In particular, we see that well-chosen collections $\mathcal{C}$ that track curves accumulate, up to rescaling, on unique distributions on a curve. Thus, to get curve-first results for forced motion, we need to exert enough control on the collection $\mathcal{C}$ to specify a limit up to overall scaling. And to see how to exercise that control, we can investigate the character of the distributional results.

For instance, in the case of a charged body, requiring a distributional $\mathbf{T}^{ab}$, supported on a curve γ, to satisfy the dominant energy condition implies that the charge-current density must be, at most, order one. As noted in Sect. 2, one can give a complete treatment of this case; when one does so, one finds contributions to

[22] Consider this result in connection with the remarks in footnote 10: as noted there, the dominant energy condition for distributions is incompatible with higher order distributions, and thus, with point particles carrying non-vanishing angular momentum. Here we see an even stronger result, which is that, in the small body limit, extended bodies all satisfying the dominant energy condition must have vanishing angular momentum (per unit mass).

the motion of the body arising from dipole moments of the charge-current density. Our only hope of getting a unique distributional limit, then, is if we can somehow control, on the way to the limit, what the contributions from the electric and magnetic dipole moments will be—for instance, by requiring that they be suitably bounded, in the limit, by the mass of the body.

We make this idea precise as follows. Let $\mathcal{C}$ be a collection of pairs (T^{ab}, J^a) of smooth fields, where each T^{ab} satisfies the dominant energy condition. We will say that a number $\kappa > 0$ bounds the charge-to-mass ratio of the elements of $\mathcal{C}$ if, for any unit timelike vector t^a at a point, and any pair $(T^{ab}, J^a) \in \mathcal{C}$,

$$|J^a t_a| \leq \kappa T^{ab} t_a t_b.$$

This condition captures the idea that the collection $\mathcal{C}$ does not include elements whose charge density relative to any observer becomes arbitrarily large, relative to its mass density. Since in the small body limit, a "dipole moment" looks like a charge density that goes from infinitely large (and positive) to infinitely large (and negative) over a vanishingly small region, bounding the charge density in this way forces contributions from dipole moments to vanish in the small body limit. On a more technical level, since we know that T^{ab} exhibits order zero behavior in the small body limit (by virtue of the energy condition), bounding J^a by T^{ab} enforces order zero behavior on J^a as well.

We then get the following result.

Theorem 8 *Let (M, g_{ab}) be a spacetime, F_{ab} an antisymmetric tensor field on M, and γ a timelike curve. Let $\mathcal{C}$ be a collection of pairs, (T^{ab}, J^a), of tensor fields on M, where each T^{ab} satisfies the dominant energy condition, each J^a satisfies $\nabla_a J^a = 0$, and each pair satisfies $\nabla_b T^{ab} = F^a{}_b J^b$. Suppose the collection has charge-mass ratio bounded by $\kappa \geq 0$ and that it tracks γ. Then there exists a sequence of pairs, $(\overset{n}{T}{}^{ab}, \overset{n}{J}{}^a)$, each a multiple of some element of $\mathcal{C}$, that converge, as distributions, to $(u^a u^b \delta_\gamma, \kappa' u^a \delta_\gamma)$, for some number κ' satisfying $|\kappa'| \leq \kappa$.*

Corollary 9 *The curve γ is a Lorentz force curve with charge-to-mass ratio κ'.*

This result captures a sense in which the Lorentz force law is a theorem of electromagnetism—and it also shows that this theorem has the same "curve-first" character as, say, the Geroch–Jang theorem. Note that a crucial assumption in Theorem 8 is that for each pair (T^{ab}, J^a) in $\mathcal{C}$, $\nabla_b T^{ab} = F^a{}_b J^b$—just as, in Theorem 4, a crucial assumption is that $\nabla_b T^{ab} = \mathbf{0}$.[23]

[23]Note, too, that the subtleties regarding the status of the conservation condition discussed in Weatherall (2011b, 2019) arise here, too: in particular, although for *sources* to Maxwell's equation, $\nabla_b T^{ab} = F^a{}_b J^b$ holds automatically, as a consequence of Maxwell's equations (just as $\nabla_b T^{ab} = \mathbf{0}$ holds for sources in Einstein's equation), here we are considering *test* matter in Maxwell's equations, since the background field F_{ab} is fixed in advance. One could imagine considering a variation of this result, along the lines of the Ehlers–Geroch theorem, that allows electromagnetic backreaction, or even that allows both electromagnetic and gravitational backreaction. Though I

So we see that Theorem 4 and its corollaries substantially strengthen the consequent of curve-first results—by giving the universal limiting behavior of certain sequences of smooth fields—and in doing so, provides hints about how to extend these results to forced motion. This is one way in which they extend curve-first results. The second way is that they weaken the premises. In particular, they permit matter to be non-vanishing far from γ, as long as the quantity of such matter can be made arbitrarily small in any particular region. Hence, these results apply to solutions of hyperbolic systems, such as Maxwell's equations and the Klein–Gordon equation. The basic idea is that the solutions to a hyperbolic system—say, Maxwell's equations—naturally give rise to a collection $\mathcal{C}$ of smooth fields T^{ab}. Insofar as these collections satisfy the dominant energy condition and are divergence-free, we can then apply the theorems above.

More precisely, fix a globally hyperbolic spacetime (M, g_{ab}), and let $\mathcal{C}$ be the collection of energy-momentum tensors associated with solutions of the source-free Maxwell equations on that spacetime.[24] It immediately follows, as a consequence of Maxwell's equations themselves, that each element of $\mathcal{C}$ is divergence-free and satisfies the dominant energy condition.[25] We can thus apply Theorem 6 to conclude that $\mathcal{C}$ can track *only* timelike and null curves, and apply Theorem 4 to conclude that if it tracks any curves at all, they must be geodesics.

We cannot, however, conclude from the general analysis that $\mathcal{C}$ tracks any curves at all. For that, we need to analyze the solutions to Maxwell's equations.[26] In fact, we find that $\mathcal{C}$ tracks all null geodesics; it tracks no timelike geodesics. It follows that there exist sequences of electromagnetic fields whose energy-momentum tensors converge to multiples of a δ distribution supported on null geodesics. This captures the sense in which light rays follow null geodesics, and it makes the so-called optical limit of electromagnetism a special case of more general theorems concerning small body motion. Note, however, that we have not avoided the sort of reasoning that goes into the optical limit altogether—the fact that one can form long-lasting wave packets with high-frequency solutions to Maxwell's equations is essential to the argument that $\mathcal{C}$ tracks any curves at all.

It is important to emphasize how this approach has avoided the problem with distributional solutions to hyperbolic systems described in Sect. 2: we do not require the electromagnetic fields themselves to converge to any distribution, and so we do not claim that the limiting distribution $\mathbf{T}^{ab}$ is the energy-momentum distribution associated with any particular solution. Rather, we claim that the limiting distribution approximates the energy-momentum properties of real solutions that are

do not know of any technical barriers to such results, formulating them is a delicate matter and we have not pursued it.

[24]We require that the spacetime be globally hyperbolic so that we are certain to have "enough" solutions to Maxwell's equations; one could imagine relaxing this requirement.

[25]See Malament (2012b, §2.6) for a discussion of this point.

[26]The relevant arguments concerning Maxwell's equations, and the other equations discussed below, are given in Geroch and Weatherall (2018, §4).

concentrated near a curve, without having any "underlying" field associated with it.

I will conclude this section by briefly discussing one more example, because it has some unexpected features. Consider the collection $\mathcal{C}$ of energy-momentum tensors associated with solutions of the mass m Klein–Gordon equation on our spacetime (M, g_{ab}). As with Maxwell's equations, each element of $\mathcal{C}$ is divergence-free and satisfies the dominant energy condition, and so once again $\mathcal{C}$ can track *only* timelike and null geodesics. In fact, one can show that $\mathcal{C}$ tracks all null geodesics; it tracks no timelike geodesics.

This result is perhaps surprising: after all, one might expect mass $m > 0$ Klein–Gordon fields to be *massive*, i.e., to give rise, in the small body limit, to massive particles, following timelike, not null, geodesics. The reason this does not happen turns on an ambiguity in the meaning of "mass." The parameter m in the Klein–Gordon equation does, in a certain sense, characterize the mass of the particle. But given a solution to the Klein–Gordon equation, m is neither the mass density associated with the solution at any point, nor is it the "total mass" associated with any spacelike slice (if suitable slices even exist).[27] And if we are thinking of the "particle" that arises in the small body limit of solutions to the Klein–Gordon equation, we should generally expect, for *any* fixed mass $m > 0$, that as the spatial support of the body approaches a curve, the "total mass" of the body, i.e., the integrated mass on a suitable spacelike slice, will approach zero. Thus, for any fixed m, we should think of small body Klein–Gordon solutions as behaving like massless particles. Another way to see the same conclusion is that, if one imagines trying to make a Klein–Gordon wave packet propagate more and more tightly along a curve, one needs to move to higher and higher frequency solutions. But these correspond to higher and higher velocities for the "massive" particle one is trying to construct, and ultimately converge to a null geodesic.

If we want to consider particles that are "massive", even in the limit, then, we need to consider not solutions the Klein-Gordon equation for fixed m, but rather solutions of the mass m Klein-Gordon equation *for all* $m > 0$. With this modification, we find that the collection $\mathcal{C}$ of energy-momentum tensors associated with all such solutions tracks all timelike and null geodesics.

Finally, I remark that one can also consider charged Klein–Gordon fields with a fixed background electromagnetic field; in this case, one can construct a collection $\mathcal{C}$ of pairs of energy-momentum tensors and charge-current densities for all solutions to Klein–Gordon equations with $m > 0$ and fixed charge-to-mass ratio κ. This collection will satisfy the conditions of Theorem 8, and so these fields can track only Lorentz force curves (and null curves).

[27] In what follows, when I write of "total mass," readers who are troubled by this notion should suppose we are in Minkowski spacetime, or a suitable asymptotically flat spacetime, where such notions make sense.

4 Dynamics, Inertia, and Spacetime Geometry

As I noted in the introduction, inertial structure, encapsulated by the geodesic principle, provides a powerful link between motion and physical geometry. It identifies a geometrically privileged class of curves with a physically privileged class of motions—hence giving physical significance to the notion of "geodesy." This result also has a converse, which I did not mention above: all metric geometry is encoded in the class of inertial trajectories. In particular, a classic result due to Weyl (1922) establishes that if two Lorentzian metrics agree on all null and timelike geodesics, up to reparameterization, then they are constant multiples of one another (Malament, 2012b, Prop. 2.1.4).

But the geodesic principle concerns point particles, and as I argued in Sect. 2, the status of such objects is unclear in general relativity. This puts some pressure on the foundational significance of the geodesic principle—and on the link between geometry and motion that it provides. What should we make of a foundational principle that, by the lights of the theory of which it is part, relies on the counterfactual behavior of impossible objects?[28]

Fortunately, there is another way. The methods described in Sect. 3 provide the resources to capture the link between motion and physical geometry directly via the solutions to matter field equations (i.e., hyperbolic systems), without any reference to point particles. The key idea is, once again, tracking, which allows us to state a new form of the geodesic principle as follows: The energy-momentum tensors associated with solutions to source-free matter field equations track (only) timelike or null geodesics.

What does this formulation express? First, it once again captures something about inertial, i.e., force-free, motion. This is because we restrict attention to *source-free* fields, where we understand sources to be interactions with other forms of matter.[29] It is these solutions that one would expect to be associated with divergence-free energy-momentum tensors. It also associates certain force-free motions of physical bodies with a geometrically privileged class of curves. Now, though, that association runs via a particular limiting construction, concerning the curves near which solutions to these equations can be made to propagate. It tells us something about how the solutions to these equations behave.

[28]One might respond that the Geroch–Jang and Ehlers–Geroch theorems do not explicitly refer to point particles, and so these, too, permit one to state the geodesic principle without reference to point particles. Fair enough. But from my perspective the main appeal of the current proposal is precisely that it is an assertion *about field equations*, and as we have seen, this is precisely what one cannot get from the Geroch–Jang and Ehlers–Geroch constructions. I am grateful to an anonymous referee for raising this objection.

[29]There is an interesting question lurking in the background here, which is: can we always unambiguously identify "source terms" in a differential equation? In standard cases in physics, it is generally clear what counts as a source. But I will not attempt to give an analysis of this concept here, and will proceed on the assumption that it is sufficiently clear for current purposes.

Remarkably, in this new form the geodesic principle is (almost) a theorem *as stated*. The results in Sect. 3 establish that it holds for a system of field equations whenever the energy-momentum tensors associated with source-free solutions have two properties: (1) they are divergence-free with respect to the spacetime derviative operator ∇ and (2) they satisfy the dominant energy condition.

The first of these conditions holds in considerable generality for matter whose dynamics can be derived from a Lagrangian density in a certain standard way.[30] In particular, consider some species of matter Φ^X in a spacetime (M, g_{ab}). Suppose the dynamics for Φ^X follow from extremizing an action $I[\Phi^X, g_{ab}] = \int_M \mathcal{L}(\Phi^X, g_{ab})$ depending only on g_{ab}, Φ^X, and its covariant derivatives. If Φ^X_0 is a solution to the resulting equations in (M, g_{ab}), then $T^{ab} := \left(\frac{\delta\mathcal{L}}{\delta g_{ab}}\right)_{|\Phi^X_0, g_{ab}}$ is divergence-free with respect to the derivative operator compatible with g_{ab}. Hence, for a broad class of matter that includes all candidates for "fundamental" matter fields in general relativity, energy-momentum is conserved relative to the metric appearing in its dynamics—i.e., the one determining the notions of length, duration, and angle salient to its evolution.

But what about the energy condition?[31] First, we remark that an energy condition is essential to the arguments given in Sect. 3. In particular, the dominant energy condition plays two roles there. First, it enforces "positivity." The basic idea behind tracking is to use the fact that we have a class of tensor fields whose action on a certain class of test fields is always non-negative, to "measure" the energy-momentum in different regions. This, recall, is how we capture the idea that there is "more" T^{ab} in a region near a curve than there is far from the curve. If we tried to drop the energy condition all together, tracking would no longer make sense, because the fields T^{ab} under consideration would not necessarily be positive when acting on any particular set of test fields. On the other hand, it is likely that a weaker energy condition would suffice in this role. The key seems to be to require that all fields T^{ab} in a collection $\mathcal{C}$ lie, at each point, within some convex cone.

The second role that the energy condition plays is that it enforces causality. That is, the dominant energy condition is what rules out the possibility of collections tracking *spacelike* geodesics, as in Theorem 6. It does not appear to be the case that weaker energy conditions could suffice for this role.[32] Thus, it seems to be

---

[30]This claim is well-known and widely discussed in the physics literature; see, for instance, Wald (1984, Appendix E) for an argument. For further discussion in a foundational context, with particular emphasis on the relationship between this claim and the geodesic principle in general relativity and other theories, see Weatherall (2019).

[31]For a discussion of the status of energy conditions in general relativity, see Curiel (2017).

[32]In effect, this is what is shown in Weatherall (2012). Note, however, that the *strengthened* dominant energy condition considered there, which is necessary for the Geroch–Jang theorem as stated, would not be natural in the current context. The reason is that distributions do not take values at points, and so requiring that they have certain behavior at points where they are non-vanishing is awkward to express. At best one would have to recast the condition in terms of the support of the distribution.

the case that one could relax the energy condition, and still conclude that solutions to a field equation track only geodesics. But the full claim that a collection of T^{ab} fields tracks only timelike or null geodesics apparently requires at least the dominant energy condition.

So we need the dominant energy condition. Fortunately, it holds automatically for many fields of physical interest. For instance, the dominant energy condition always holds for the energy-momentum tensors associated with source-free solutions to Maxwell's equations, for solutions to the (non-negative mass) Klein–Gordon equation, and so on. But it does not hold for *all* equations that one might be interested in. In particular, solutions to the Dirac equation may not satisfy even the weak energy condition.[33] This suggests that it is the dominant energy condition that is key to whether a given form of matter, with dynamics derivable from a suitable Lagrangian, will satisfy the (new) geodesic principle—and also that it is not clear that all matter fields of physical interest *do* satisfy the new geodesic principle.[34]

This discussion suggests that the status of the dominant energy condition deserves more attention. In particular, one would like to identify the conditions under which the energy-momentum tensor associated with solutions to a given matter field equation are certain to satisfy the energy condition. Of special interest would be to articulate the relationship between the dominant energy condition, on the one hand, and the "causal cone" associated with a hyperbolic system, which captures a (different) sense in which solutions to a system of equations may propagate causally.[35]

Acknowledgments I am grateful to Harvey Brown, Bob Geroch, Dennis Lehmkuhl, David Malament, and Bob Wald for many helpful conversations related to this material. I am also grateful to audiences at the Eighth Quadrennial International Fellows Conference (2016) in Lund, Sweden; New Directions in Philosophy of Physics (2017) in Tarquinia, Italy; the Pacific Institute of Theoretical Physics at the University of British Columbia; the workshop "Thinking about Space and Time" at the University of Bern; the Max Planck Institute for Gravitational Physics in Potsdam, Germany; the Black Hole Initiative at Harvard University; the Einstein Papers Project at Caltech; and Utrecht University, and especially to Lars Andersson, Joshua Goldberg, Jim Hartle, Bob Wald, and Shing-Tung Yau for probing questions, comments, and objections.

[33]Observe that this failure to satisfy the energy conditions is not obviously related to the fact that Dirac fields have "intrinsic" angular momentum (though it is related to the fact that they are spinors). (Recall fns. 10 and 22.)

[34]One might worry that this last observation is a problem for the proposed formulation of the geodesic principle in terms of tracking. But I do not think there is a real concern. Source-free matter that tracks non-geodesic curves is every bit as much a problem for other formulations of the geodesic principle as the present one—and at least on the proposed formulation, the tension between such matter and geodesic motion is immediately manifest.

[35]There has been some discussion of this relationship in both the physics and philosophy literatures (Geroch, 1996; Earman, 2014; Weatherall, 2014; Wong, 2011), but it does not seem that a fully satisfactory answer is available.

References

Asada, H., Futamase, T., Hogan, P. A., 2011. Equations of motion in general relativity. Oxford University Press, Oxford, UK.

Brown, H. R., 2005. Physical Relativity. Oxford University Press, New York.

Colombeau, J. F., 2000. New generalized functions and multiplication of distributions. Vol. 84. Elsevier.

Curiel, E., 2017. A primer on energy conditions. In: Lehmkuhl, D., Schiemann, G., Scholz, E. (Eds.), Towards a Theory of Spacetime Theories. Birkhäuser, Boston, MA, pp. 43–104.

D'Eath, P. D., 1975. Interaction of two black holes in the slow-motion limit. Physical Review D 12 (8), 2183.

Earman, J., 2014. No superluminal propagation for classical relativistic and relativistic quantum fields. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 48, 102–108.

Ehlers, J., Geroch, R., 2004. Equation of motion of small bodies in relativity. Annals of Physics 309, 232–236.

Einstein, A., Grommer, J., 1927. Allgemeine Relativitätstheorie und Bewegungsgesetz. Verlag der Akademie der Wissenschaften, Berlin.

Geroch, R., 1996. Partial differential equations of physics, arXiv:gr-qc/9602055.

Geroch, R., 2011. Faster than light? In: Plaue, M., Rendall, A., Scherfner, M. (Eds.), Advances in Lorentzian Geometry. American Mathematical Society, Providence, RI, pp. 59–80.

Geroch, R., Jang, P. S., 1975. Motion of a body in general relativity. Journal of Mathematical Physics 16 (1), 65.

Geroch, R., Traschen, J., 1987. Strings and other distributional sources in general relativity. Physical Review D 36 (4), 1017.

Geroch, R., Weatherall, J. O., 2018. The motion of small bodies in space-time. Communications in Mathematical Physics (forthcoming). https://doi.org/10.1007/s00220-018-3268-8.

Gralla, S. E., Harte, A. I., Wald, R. M., 2009. Rigorous derivation of electromagnetic self-force. Physical Review D 80 (2), 024031.

Gralla, S. E., Wald, R. M., 2011. A rigorous derivation of gravitational self-force. Classical and Quantum Gravity 28 (15), 159501.

Grosser, M., Kunzinger, M., Oberguggenberger, M., Steinbauer, R., 2001. Geometric theory of generalized functions with applications to general relativity. Vol. 537. Springer Science & Business Media, Dordrecht.

Lehmkuhl, D., 2017a. General relativity as a hybrid theory: The genesis of Einstein's work on the problem of motion. Studies in History and Philosophy of Modern Physics (forthcoming). https://doi.org/10.1016/j.shpsb.2017.09.006.

Lehmkuhl, D., 2017b. Literal versus careful interpretations of scientific theories: The vacuum approach to the problem of motion in general relativity. Philosophy of Science 84 (5), 1202–1214.

Malament, D., 2012a. A remark about the "geodesic principle" in general relativity. In: Frappier, M., Brown, D. H., DiSalle, R. (Eds.), Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos. Springer, New York, pp. 245–252.

Malament, D. B., 2012b. Topics in the Foundations of General Relativity and Newtonian Gravitation Theory. University of Chicago Press, Chicago.

Mathisson, M., 1931. Die mechanik des materieteilchens in der allgemeinen relativitätstheorie". Zeitschrift für Physik 67, 826–844.

Mino, Y., Sasaki, M., Tanaka, T., 1997. Gravitational radiation reaction to a particle motion. Physical Review D 55 (6), 3457.

Papapetrou, A., 1951. Spinning test-particles in general relativity. i. Proc. R. Soc. Lond. A 209 (1097), 248–258.

Poisson, E., Pound, A., Vega, I., 2011. The motion of point particles in curved spacetime. Living Reviews in Relativity 14 (7).

Puetzfeld, D., Lämmerzahl, C., Schutz, B. (Eds.), 2015. Equations of motion in relativistic gravity. Springer, Heidelberg, Germany.

Samaroo, R., 2018. There is no conspiracy of inertia. British Journal for the Philosophy of Science. 69 (4), 957–982.

Souriau, J.-M., 1974. Modèle de particule à spin dans le champ électromagnétique et gravitationnel. Annales de l'Institut Henri Poincaré Sec. A 20, 315.

Steinbauer, R., Vickers, J. A., 2006. The use of generalized functions and distributions in general relativity. Classical and Quantum Gravity 23 (10), R91.

Sternberg, S., Guillemin, V., 1984. Symplectic Techniques in Physics. Cambridge University Press, Cambridge.

Sus, A., 2014. On the explanation of inertia. Journal for General Philosophy of Science 45 (2), 293–315.

Tamir, M., 2012. Proving the principle: Taking geodesic dynamics too seriously in Einstein's theory. Studies in History and Philosophy of Modern Physics 43 (2), 137–154.

Thorne, K. S., Hartle, J. B., 1985. Laws of motion and precession for black holes and other bodies. Physical Review D 31 (8), 1815.

Wald, R. M., 1984. General Relativity. University of Chicago Press, Chicago.

Weatherall, J. O., 2011a. The motion of a body in Newtonian theories. Journal of Mathematical Physics 52 (3), 032502.

Weatherall, J. O., 2011b. On the status of the geodesic principle in Newtonian and relativistic physics. Studies in the History and Philosophy of Modern Physics 42 (4), 276–281.

Weatherall, J. O., 2012. A brief remark on energy conditions and the Geroch-Jang theorem. Foundations of Physics 42 (2), 209–214.

Weatherall, J. O., 2014. Against dogma: On superluminal propagation in classical electromagnetism. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 48, 109–123.

Weatherall, J. O., 2017. Inertial motion, explanation, and the foundations of classical spacetime theories. In: Lehmkuhl, D., Schiemann, G., Scholz, E. (Eds.), Towards a Theory of Spacetime Theories. Birkhäuser, Boston, MA, pp. 13–42.

Weatherall, J. O., 2019. Conservation, inertia, and spacetime geometry. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 67, 144–159.

Weyl, H., 1922. Space–Time–Matter. Methuen & Co., London, UK, reprinted in 1952 by Dover Publications.

Wong, W. W.-Y., 2011. Regular hyperbolicity, dominant energy condition and causality for Lagrangian theories of maps. Classical and Quantum Gravity 28 (21), 215008.

The Metaphysics of Machian Frame-Dragging

Antonio Vassallo and Carl Hoefer

Abstract The paper investigates the kind of dependence relation that best portrays Machian frame-dragging in general relativity. The question is tricky because frame-dragging relates local inertial frames to distant distributions of matter in a time-independent way, thus establishing some sort of nonlocal link between the two. For this reason, a plain causal interpretation of frame-dragging faces huge challenges. The paper will shed light on the issue by using a generalized structural equation model analysis in terms of manipulationist counterfactuals recently applied in the context of metaphysical enquiry by Schaffer (Philos Stud 173:49–100, 2016) and Wilson (Metaphysical causation. Noûs 2017). The verdict of the analysis will be that frame-dragging is best understood in terms of a novel type of dependence relation that is half-way between causation and grounding.

1 Introduction

It is virtually impossible to address the problem of the origin of inertia in spacetime theories without mentioning Ernst Mach and his "The Science of Mechanics" (Mach, 1883). Indeed, Mach's views on inertia have been discussed and analyzed at length in the philosophical literature. The consensus among Mach's commenters is that his discomfort with the origin of inertia in Newtonian mechanics comes from epistemological considerations filtered through empiricist inclinations. For Mach, it is highly unsatisfying to link inertia to Newtonian absolute space for the simple reason that such a space is unobservable. If the aim of physics—and science in general—is to provide a picture of the world based on experience, then there is no place in this picture for elements that elude experience in one way or another.

A. Vassallo (✉)
International Center for Formal Ontology, Warsaw University of Technology, Warsaw, Poland

C. Hoefer
Universitat de Barcelona, Gran Via de les Corts Catalanes, Barcelona, Spain

ICREA, Barcelona, Spain

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_11

At this point, however, the consensus breaks down as to how much further Mach pushed these considerations. On a prudent reading, it seems that Mach simply argues for a mere redescription of the role of inertia in classical mechanics, that is, that inertia should not be referred to absolute space, but to a suitably defined reference frame (e.g., in the case of Newton's bucket experiment, the inertial effects arising should be referred to the distant background of fixed stars). A more radical reading instead suggests that Mach has in mind not just a mere redescription of classical mechanics, but a brand new physical theory in which inertia originates from the overall distribution of masses in the universe (see Huggett and Hoefer, 2017, section 8). It is no secret that Einstein saw this "radical Mach" as one of his main inspirations in his quest for general relativity (GR). Although there are still disputes about whether GR fully complies with a radical interpretation of Mach's views (see, for example, Barbour and Pfister, 1995, chapter 3), it is a well-established fact that, under certain physical conditions, the theory predicts that local inertial frames are determined to a certain degree by the surrounding material distribution. This is the case for frame-dragging effects, such as the Einstein–Lense–Thirring effect, according to which a gyroscope orbiting a huge rotating mass distribution will show a precession that is directly related to the angular momentum of the distribution (Pfister, 2007).

In a nutshell, frame-dragging effects imply that the axes of a local inertial frame are not fixed and independent of the surrounding material distribution, which in turn means there is no such thing as an absolute space acting as a fixed reference for an inertial compass. Section 2 will provide a very brief introduction to rotational frame-dragging effects in GR, from modest Machian effects implying that local inertial frames are determined by both the state of motion of surrounding matter distribution and the condition of asymptotic flatness at infinity, to fully Machian effects dictating that inertial frames are defined with respect to the overall matter distribution in the universe. The key point to be highlighted is that, in any case, the determination relation is established in a synchronic way. This apparently means that local inertial frames would instantaneously "feel" any change in the surroundings' state of motion, irrespective of how far away in space such a change is triggered. This may be a source of unease for those inclined to consider frame-dragging as a result of some sort of physical interaction. In fact, the absence of a retarded mechanism underlying these effects might lead them to consider frame-dragging as an instance of action-at-a-distance. This would in turn raise delicate questions for metaphysicians who want to read a causal dependence off from the frame-dragging mechanism, for they would then apparently be forced to accept some kind of superluminal causation. Although this would not by itself be a fatal blow to a causal interpretation of frame-dragging (after all, one might simply take it to show that in GR there is superluminal causation), still one might take this controversial issue as a motivation to pursue an alternative metaphysical analysis. This is exactly what is done in Sect. 3, where the framework of structural equation models (SEM) will be discussed. The framework helps to analyze the dependency relations underlying a set of correlations by (1) constructing a mathematical model of such dependencies and (2) counterfactually testing the model in a manipulationist fashion. This tool has been so far used mainly in social science to extract causal information from huge datasets but, recently, some metaphysicians have noted that the framework works in principle with any kind of determination relation, and in particular with grounding. The aim of this paper is to apply the SEM framework to shed light on the issue of finding the "right" dependence relation depicted by frame-dragging: the results of our analysis are presented in Sect. 4 and further discussed in Sect. 5.

2 Rotational Frame-Dragging Effects in GR

The conceptual path that leads to general relativistic frame-dragging can be traced back to Newton's bucket argument. This thought experiment was meant to show that the relative rotation of water with respect to the bucket's walls was neither necessary nor sufficient to ground the centrifugal forces responsible for the concavity of the water's surface. Newton's conclusion was that centrifugal forces arose as a result of water rotating with respect to absolute space, since all possible inertial frames were tied to this latter entity. Mach famously replied to this challenge by pointing out that all we can *observe* is that the curving of the water's surface is not determined by its state of motion with respect to the walls of the bucket, but is determined by its state of motion relative to the frame of the fixed stars. We do not know, Mach went on to say, whether the curving of the water's surface would be just the same if all the matter of the fixed stars were removed from the universe, nor whether a curving effect might in fact be produced by mere relative rotation with respect to the walls of a bucket, if the latter were "several leagues thick". GR vindicates Mach's point of view on dragging effects in that (a) in asymptotically flat universes any local inertial compass would nonetheless feel the presence of a surrounding rotating matter distribution, and (b) in nonasymptotically flat universes inertial frames appear to be, in some models at least, completely determined by the overall mass distribution. With these facts in mind, we can say that dragging effects in GR come in different degrees of Machianity. In the following, we will review the most important results, placing them in an ascending scale of Machianity from (a) to (b), without any pretense to be exhaustive or mathematically rigorous.

The less Machian frame-dragging case was first discussed in the early years of GR, by Einstein, Lense, and Thirring (here we draw from Ciufolini and Wheeler, 1995, section 6.1 and Misner et al., 1973, sections 18.1 and 19.1). Consider a slowly rotating ideally spherical body with mass M and angular momentum $\mathbf{J}$ in a stationary, asymptotically flat spacetime. In the weak field limit, we can use linearized gravity, that is, we can decompose the metric $g_{\mu\nu}$ into a flat Minkowski background $\eta_{\mu\nu}$ plus a perturbation term $h_{\mu\nu}$ (throughout the paper we will set $G = c = 1$):

$$g_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu}, \tag{1}$$

and define:

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\left(\eta^{\alpha\beta}h_{\alpha\beta}\right). \tag{2}$$

The general solution of the linearized field equations will then be:

$$\bar{h}_{\mu\nu} = 4\int \frac{T_{\mu\nu}(\tau, \mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x', \tag{3}$$

where $\tau = t - |\mathbf{x} - \mathbf{x}'|$ is the "retarded time", which models the fact that big enough perturbations of the background would propagate from the source at a finite velocity as gravitational waves. However, in the present case, velocities are too low to produce such big perturbations. Moreover, in general, by expanding $h_{\mu\nu}$ in powers of $\frac{\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|} \equiv \frac{\mathbf{x}'}{r}$, we see that the higher-order radiation terms die out as $\frac{1}{r}$ so, if the radius of the body is big enough, no retardation effect will be seen.

In this context, the infinitesimal $0i$ $(i = 1, 2, 3)$ components of Einstein's field equations can be written in the Lorenz gauge as:

$$\delta\bar{h}_{0i} = 16\pi\rho v^i, \tag{4}$$

with ρ the density and v^i the linear 3-velocity of the mass distribution. The solution will be:

$$\bar{h}_{0i} = -4\int \frac{\rho(\mathbf{x}')v^i(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x. \tag{5}$$

In the end, by passing to spherical coordinates, the metric outside the body will approximately be (in the appropriate (Kerr) gauge, where $\mathbf{J} = (0, 0, J)$):

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2\left(d\theta^2 + sin^2\theta d\phi^2\right) - \\ -\frac{4J}{r}sin^2\theta dt d\phi. \tag{6}$$

The off-diagonal term $\bar{h}_{t\phi}$ in (6) comes from (5) and from the fact that $\mathbf{J} = \int \mathbf{x} \times (\rho\mathbf{v})d^3x$, and can be considered some sort of potential in analogy to the magnetic potential in electrodynamics. More precisely, it is a "dragging potential" because it can be shown that a gyroscope orbiting the massive body would precess—with respect to an observer at infinity—with angular velocity:

$$\mathbf{\Omega} = -\frac{1}{2}\nabla \times \mathbf{A}, \tag{7}$$

where $\mathbf{A} = \left(0, 0, -\frac{2J}{r}sin^2\theta\right)$ and ∇ is the covariant spatial derivative. Equation (7) is thus a gravitational analogue of Faraday's law of induction in electrodynamics.

Without entering into further mathematical detail, we already see what is Machian about the result (7): since the gyroscope determines the axes of a local inertial frame, the state of motion of the mass distribution influences such a determination in the vicinity of its surface. Note that $\mathbf{\Omega}$ varies considerably at different locations. For example, near the poles the inertial frames rotate in the same direction as the massive body, while near the equator the inertial frames rotate in the opposite direction. We also see what is *not* Machian here (basically, everything else): the rotation of the massive body and the precession of the gyroscope are referenced to the *infinity* fixed ab initio without reference to any mass distribution. Formally, this means that (7) is solved modulo a constant of integration that the asymptotic flatness condition fixes to be $\mathbf{\Omega}(\infty) = 0$. In this sense, for $r \to \infty$ we get the "true" inertial frame, which suspiciously looks like an absolute space in the Newtonian sense.

A slightly more Machian dragging effect arises if we consider a slowly rotating massive spherical shell. To determine the metric inside this body, we just note that, in the case $\mathbf{J} = 0$, this metric would be the flat Minkowskian one $\eta_{\mu\nu}$. Hence, for $\mathbf{J} \neq 0$, in the weak field limit we can apply again (1), $h_{\mu\nu}$ again being a rotational perturbation small enough that the resulting metric is stationary (in order to avoid higher order contributions from gravitational waves). With this machinery in place, we can apply the same reasoning above, thus finding that the axes of the gyroscope near the center of the shell would be dragged with respect to infinity. This situation is slightly more Machian than the previous one because, in the reference frame of the gyroscope, the metric inside the shell appears flat, thus hinting at the fact that the matter distribution defines "its own" inertial frames inside the shell. However, the Machianity of the situation ends here, because the rotations involved are not defined relationally but refer, again, to infinity. Furthermore, the approximations used in the model make it too much of a toy model. For example, we need to impose physically implausible conditions on the stress-energy tensor in order to keep the radius of the shell constant, this in turn makes it troublesome to "connect" the metric inside the shell with that on the outside.

More physically realistic models of slowly rotating mass shells with a time dependent radius (in order to model their expansion/contraction) are due to Brill and Cohen (1966), Lindblom and Brill (1974), and Pfister and Braun (1985). Roughly, these authors apply a slightly more sophisticated perturbation approach to the outside of the shell. Given that, by Birkhoff's theorem, the metric outside a static spherical distribution of matter in an otherwise empty universe is the Schwarzschild one, they start by a rotationally perturbed version of it, which can be generically written as:

$$ds^2 = e^{2\alpha}dt^2 - e^{2\beta}dr^2 - r^2\left[d\theta^2 + sin^2\theta\,(d\phi - \omega dt)^2\right], \tag{8}$$

where $\omega = 2\frac{J}{r^3}$, and $e^{2\alpha}$ and $e^{2\beta}$ are appropriate functions of the shell's mass determined by the field equations plus the state equation for matter. The metric inside the shell is taken to be flat, although written in rotating coordinates depending also on the radius of the shell. In this way, the authors are able to give a full description of the system, including the connecting region on the shell. The angular momentum $\mathbf{J} = (0, 0, J)$ within a region r depends on the $t\phi$ component of the stress-energy tensor through:

$$J = \int_0^r \int \int T^t_\phi \sqrt{-g} d\theta d\phi dr. \tag{9}$$

Again, we do not need to go too much into technicalities here: the interesting point for us is that all these authors recover frame-dragging results that are a generalized form of (7):

$$e^{-\alpha}\mathbf{\Omega} = -\frac{1}{2}e^{\alpha}\nabla \times \mathbf{A}, \tag{10}$$

where the factor $e^{-\alpha}$ accounts for the increase of rotation rate of the shell due to the gravitational slowing of the clocks close to the shell with respect to those at infinity.

Lynden-Bell et al. (1995) further generalize these results by considering N nested freely falling shells, the metric between any two shells being of the form (8), and taking the continuous limit (i.e., $N \to \infty$). They find the interior variation of the metric's perturbation ω to be:

$$\frac{\partial \omega}{\partial r} = -6e^{\alpha+\beta}\frac{J}{r^4}. \tag{11}$$

If we assume asymptotic infinity, so that $\omega \to 0$ for $r \to \infty$, we can integrate (11) to get the overall perturbation distribution over space:

$$\omega = 2\left(\frac{W}{r^3}J + \int_r^\infty \frac{W}{r'^3}\frac{\partial J}{\partial r'}dr'\right), \tag{12}$$

where W is an appropriate weight function that depends on $e^{\alpha+\beta}$ and r. Roughly, (12) is a measure of how much dragging comes from any spatial region of the model. Hence, this model nicely describes a case where local inertial frames are partially determined by the overall matter distribution in the universe. This is the most Machian scenario still falling in the category (a).

The subsequent step is to render the model even more Machian by assuming that there is no infinity region, i.e., by imposing that the universe is closed. In this case, it is more useful to define a new coordinate $0 < \chi < \pi$ such that $sin\chi = \frac{r}{r_{max}}$, r_{max} being the maximum size of the universe. Simple geometrical considerations lead to a new version of (11) appropriate for the closed case:

$$\frac{\partial \omega}{\partial \chi} = -6e^{\alpha+\beta}\frac{\partial r}{\partial \chi}\frac{J}{r^4}. \tag{13}$$

When integrating (13) in order to get the closed universe equivalent of (12), we have to keep in mind that there is no more boundary condition $\omega(\infty) = 0$ because there is no infinity. The only thing we can do is to integrate with respect to an arbitrarily chosen point χ_*:

$$\omega - \omega_* = 2\left(\frac{W_*}{r^3}J + \int_{\chi}^{\chi_*} \frac{W_*}{r'^3}\frac{\partial J}{\partial \chi'}d\chi'\right). \tag{14}$$

Equation (14) tells us that, since there is no privileged inertial frame at infinity, the angular momentum distribution just contains information about the *relative* rotations of inertial frames at different points. The icing on the cake is given by a mathematical result proven by the authors in this context, which states that the total angular momentum of a closed universe is necessarily zero. This is the most Machian setting among those considered so far, in which dragging effects arise. Equation (14) comes very close to Mach's idea encompassed in his reply to Newton, i.e., that the concavity on the surface of a rotating mass of water was just due to the dragging of the water's inertial frame by the relative counter-rotation of the surrounding matter in the entire universe.

These results can be further generalized by considering small perturbations of a FLRW metric:

$$ds^2 = dt^2 - a^2(t)\left[\frac{1}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta d\phi^2\right)\right], \tag{15}$$

$a(t)$ being a scale factor and k=-1,0,1 representing the (constant) spatial curvature. The perturbed metric $\tilde{ds}^2$ will feature off-diagonal elements corresponding to such perturbations. The key point to consider is that, when we want to "project" $\tilde{ds}^2$ on the unperturbed background ds^2, we have a number of gauge degrees of freedom associated with the background's underlying symmetries. In this particular case, the unperturbed metric exhibits spatial homogeneity and rotational symmetry, which, in particular, means that the associated Killing vectors are 3-dimensional. In this context, Noether's theorem implies that any conserved quantity at a given instant of cosmic time t features—at the first perturbational order—only the $0i$ $(i = 1, 2, 3)$ components of the stress-energy tensor and, hence, only the "spatial constraints" of the Einstein's field equations are involved. From all of this, we can calculate the off-diagonal h_{0i} terms in $\tilde{ds}^2$ as functions of 3-Killing displacements ξ_{ai} $(a = 1, \ldots, 6)$ by integrating the set of equations:

$$\delta h_{0i} = f^a \xi_{ai}. \tag{16}$$

Equation (16) is a general form that includes (12) and (14) as particular cases, f^a being a six-component integration constant. Note that no time-like derivative appears in this relation. In the case of an open universe, f^a is fixed by the boundary conditions at infinity (as in the case $\omega(\infty) = 0$ in (12)) but, if the universe is closed, some freedom will remain as to choose f^a, meaning that only relative motions are definable, in the same Machian spirit as (14). This latter result has been later extended—albeit in very particular gauges—to *any* FLRW cosmology, thus entirely fulfilling Mach's ideas (see Bičák et al., 2007).

To sum up, rotational frame-dragging effects are perhaps best understood if we bring in again the analogy with electrodynamics, which is apparent in (7). Such effects are induced by a dragging potential encoded in the off-diagonal terms representing small perturbations of a background metric (be it Minkowski, Schwarzschild, or FLRW). The very general form of these perturbations is (16), whose right-hand side is determined by the mass-energy distribution in the region of interest (as, for example, in (4)). In all these cases, however, the magnitude of δh_{0i} is so small that no retarded action of the form (3) will arise because no substantial gravitational radiation that literally carries the perturbation will be produced. This is an obvious disanalogy with the electrodynamic case, where physically realistic potentials—especially in long-range interactions—are retarded. The immediate consequence of this lack of retardation in the general relativistic context is that introducing any δ-sized change in the mass-energy distribution of a system will alter the magnitude of the dragging effects felt by local inertial frames *instantaneously*, that is, on the $t = \text{const.}$ surfaces. For example, if we take a spherical shell of matter centered on a point P in a FLRW universe and give it a small rotation about P, the result of the perturbation is felt instantaneously at P. As we noted above, this effect is instantaneous on a constant-t surface in the model due to the fact that its existence is derived from the spatial constraint equations, which can be thought of as one part of Einstein's field equations (note, *en passant*, that no violation of general covariance is implied, i.e., frame-dragging effects do not identify a *privileged* foliation of spacetime).

This is of course a bit of a problem—especially in a cosmological context—if we wish to say that local inertial frames are *causally* affected by the surrounding matter distribution. In fact, causation is usually understood as a diachronic relation (causes precede their effects) but, even if we are willing to relax this condition enough to admit at least synchronic causation, we are left with something that looks a bit like spooky action-at-a-distance. And even if we bite this bullet, still we will face the challenge of explaining whether, and if so why, this action amounts to—or does not amount to—some kind of superluminal physical influence.

The problems in this context are made even worse by the fact that no well-established philosophical account of causation is able to provide a neat analysis of dragging dependencies. Conserved quantity approaches à la Salmon-Dowe are notoriously ill-defined in GR, and also causal property theories such as Alexander Bird's dispositional monism are at odds with many foundational aspects of the theory (see Lam, 2010, and references therein). The situation is not better for counterfactual theories such as Lewis' or Woodward's, given that the GR's dynamical geometry gives no well-established standard to evaluate a counterfactual change against the actual situation (see Curiel, 2015, for a general discussion, and Hoefer, 2014, for the specific case of frame-dragging). The first difficulty regards a counterfactual situation involving a local change that leaves everything else untouched. In this case, even if we are working with a model of GR that admits a well-posed initial value formulation and we express this "change" as a tiny modification of the metrical and material properties in a small neighborhood on the initial space-like surface, such a change would be enough to violate the field equations, which constrain the happenings on each surface, thus making any attempt to evolve the situation to see "what would had happened if we had made this small change, leaving everything else as is in the actual world" far-fetched. The second problem regards testing counterfactuals by comparing two models of the theory, corresponding to the actual and counterfactual situations. In this case, the lack of a standard implies that the way we choose the counterfactual model is always somewhat arbitrary. This point becomes evident if we consider the counterfactual situation where all the masses in the actual world would vanish (decay into the true Higgs vacuum is a concrete possibility…). We might naturally be inclined to say that, if that happened, the Minkowski solution would be the best candidate to model the situation. However, Minkowski spacetime is not the only vacuum solution of Einstein's field equations, and no compelling mathematical argument can be made that singles out this particular solution over the others. Moreover, as is pointed out in Lynden-Bell et al. (1995), if one were to approach modelling this counterfactual situation via a series of FLRW models with ever-smaller cosmic mass, then if one starts from a closed/finite universe, as the total cosmic mass gets smaller and smaller the spacetime becomes smaller, shorter-lived, and more highly curved, eventually (in effect) *vanishing* in the limit as cosmic mass goes to zero, rather than turning into Minkowski spacetime (or any other infinite empty spacetime of GR).

The ambiguity problem affects our treatment, for example, in the Einstein–Lense–Thirring case. As a matter of fact, one can consider the external field of a spinning nearly spherical massive body in an otherwise empty universe (Kerr model) and then test the counterfactual "If the body had not spun, the local inertial frame K at a distance $|r|$ from the body's surface would not have precessed" against a Schwarzschild model for a body with same mass and radius but no angular momentum, thus finding that the counterfactual is true. In this case, our choice would be more robust because of Birkhoff's theorem. However, the problem is still there, and manifests itself in the fact that there is no objective well-established trans-model identity criterion that helps us point at K in the starting model and say that the orientation of the z-axis of the very same K in the second model does not precess. Usually, physicists make sense of such counterfactuals by stipulation: they map the perturbed space (the Kerr model, in this example) on the unperturbed background (here, the Schwarzschild model) in a way that fixes for any point P in the background its "perturbed" counterpart P'. This fixing procedure basically amounts to the condition that P and P' are always flagged with the same coordinate value. In this way, any choice of coordinates on the background will immediately fix the coordinates in the perturbed space. Thanks to this stipulation, we immediately see that we can evaluate the above counterfactual without particular worries. However, this approach to trans-model identity just works on a case-by-case basis and, therefore, cannot be adopted as a formal backbone of a general counterfactual analysis of causation.

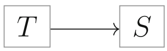
In the following section, we will consider a different approach to the analysis of causation, one which, in our opinion, is a promising framework for investigating the nature of dependence relations in GR.

3 Structural Equation Model Analysis

The umbrella title "structural equation modelling" designates a set of mathematical tools developed as early as the 1960s in parallel with the rise of database management systems. The main scope of these tools is to analyze distributions of data by best-fitting them in graph structures that explain some "mechanism" (in a loose sense) of interest underlying these distributions. So far, these models have been mainly used in sociology and psychology to, e.g., test the hypothesis that intelligence (under a certain operational measure) strongly influences academic performance. It is worth noting that this framework is tentative in nature and, as such, it progresses on a trial-and-error basis. First, a certain graph structure is proposed, which describes how data—sorted out by different kinds of variables—are related based on some explanatory hypotheses, and then it is tested, e.g., by statistical methods, how well actual data fit the network of dependencies posited: if the results are unsatisfactory, a modification of the starting hypotheses is made and the entire process is performed again.

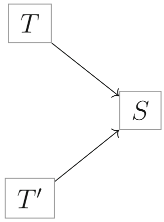
The SEM framework naturally lends itself to causal analysis and in particular to the evaluation of causal inferences (see Pearl, 2000 for one attempt to spell out this approach in detail). For our purposes, the following example will suffice (see Schaffer, 2016, section 2, for an extensive presentation and discussion of the framework in a philosophical context). Imagine we have a dataset that shows a systematic correlation between the members of two distinct samples X and Y, and we hypothesize that there is some sort of causal mechanism that makes it the case that Y depends on X. The starting point is to define an "endogenous" variable $y \in Y$ representing the dependent condition and an "exogenous" variable $x \in X$ representing the independent condition. We further require that these variables take values from an appropriate set, for example, the binary set $\{0, 1\}$, "0" meaning that the condition does not obtain and "1" meaning that it obtains. The next step is to provide a set of formal relations (e.g., equations) showing how y's value has to be evaluated on the basis of x's value. We can symbolize this relations by $y \overset{\leftarrow}{=} f(x)$, where the symbol "$\overset{\leftarrow}{=}$" makes the conjectured direction of dependence explicit. Hence, given an assignment of value a to x, the model gives the corresponding value $f(a)$ for y. The intuitive interpretation of this formal dependence is straightforward: the "wiggling" of x always (if f is deterministic) triggers a corresponding change in y in a "f" way, but the opposite does not hold. We immediately see that the best way to render this analysis of dependence in terms of wiggling is by using interventionist counterfactuals: "If an intervention on x had set its value to a, then y would have taken on the value $f(a)$" (see Woodward, 2016, for a survey of the interventionist framework). This also makes it clear why the independent variable is called exogenous: it is the one that is subjected to any process external to the system, i.e., an intervention. Finally, the model is tested against the dataset to see how accurately f captures the correlations pattern between the two samples. Although it most naturally and powerfully applies to generic or "type" causation, the method can be used to model both generic and singular ("token") cases of causation. In this way, the SEM framework permits both qualitative and quantitative causal analysis, and can be easily implemented into causal search algorithms (as in the case of Ramsey and Winberly's Tedrad project, whose codebase is freely available on GitHub). It is instructive to consider how the framework analyzes a textbook case of direct causation, that is, the shattering of a window caused by the throwing of a stone.

In this case we have two variables: "T" formalizing whether or not the fact that a (sufficiently heavy) stone is thrown obtains, and "S" formalizing whether the fact that the window shatters obtains. Both variables take values from the binary set $\{0, 1\}$. Based on everyday experience, we hypothesize that S causally depends on T and we model this dependence as $S \overset{\leftarrow}{=} T$. The graph corresponding to this model is:



The model does not need any further tweaking or supplementation to avoid an incorrect back-tracking conclusion ($S \rightarrow T$), as the Lewisian analysis of causation does, because it has the "right" counterfactual patterns already built in. As such, this framework is more parsimonious than the Lewisian one, because it dispenses with the talk of possible worlds, small miracles, and the like. It might be objected that this framework, contrary to the Lewisian spirit, is not reductive since the notion of intervention, which plays a key role in the evaluation of the causal link, presupposes an underlying causal process of the same nature as the one analyzed. While we agree that the SEM framework is nonreductive with respect to causation, we do not see it as a problem with the consistency or the reliability of the inferences drawn in this context since any intervention on T does not necessarily presuppose that we already have causal information about the relationship we want to characterize. Also, here, we can see explicitly that the model accounts equally well for the case where *a* stone is thrown through *a* window (type) and for the case where Leonardo throws a stone through Monika's window (token).

Now, imagine that there are two stone throwers instead of one. Clearly, the above model would become inaccurate and should be supplemented with a second exogenous variable T' and a second dependence equation $S \overset{\leftarrow}{=} T'$. The correct graph would thus be:



This shows how the framework easily handles cases of causal overdetermination. Similarly, the framework can account for cases of causation by the absence: roughly, the dependence of y on the absence of x is modelled by the structural equation $y \stackrel{\leftarrow}{=} 1 - x$.

It is important to stress the fact that there is no upper limit to the complexity of a model in this context. The graph can have (infinitely) many links and also a branching structure. Graphs can even be cyclic, to model situations (not uncommon in social sciences) where variables can mutually influence one another, or feedback loops exist. It is common, however, to only consider directed *acyclic* graphs (DAGs) meeting two consistency constraints: (1) that an intervention on a certain variable influences all the other variables downstream but no variable upstream, and (2) that the topology of the graph is not closed. This restriction is particularly natural when modelling token causation.

The SEM framework has recently attracted the attention of the metaphysical community because it can be easily generalized to any kind of dependence relation that is generative or directed in nature and can be modelled by a partial ordering, grounding being one such relation. Wilson (2017), shows that the analysis of all the major cases of causation can be replicated for grounding. For example, the stone/window case can be translated using the same variables and the same structural equation to the case of the existence of Socrates grounding the existence of singleton Socrates or, in general, the existence of singletons being grounded in the existence of their respective members. According to Wilson, the only difference between the stone/window case and the member/singleton one lies in the mediating principles involved, that is, the set of background conditions invoked to justify the model. In the first case but not in the second a subset of laws of nature is invoked (note how this law-based demarcation criterion does not mention or imply that causation has to be a diachronic relation). For this reason, Wilson renames causation as "nomological causation" and grounding as "metaphysical causation", going further in arguing that, in fact, standard causation and grounding are at most two species of a genus relation "causation" taken as primitive. However, in order for Wilson's unifying framework to work, two slight modifications of the original SEM scheme are in order. First of all, we need to liberalize the notion of intervention. Grounding counterfactuals involve "metaphysical interventions", some of which are impossible. Moreover, given that we accept metaphysically impossible interventions, we need to adopt a nonstandard semantics of counterpossibles which does not treat them as trivially true. For example, for the structural model $S \stackrel{\leftarrow}{=} T$

(T being whether Socrates exists, and S being whether singleton Socrates exists) to work, we have to evaluate the counterpossible "if an intervention had prevented singleton Socrates from existing, then Socrates would not have existed" as false (see Wilson, 2018).

For our purposes, we do not need to go as far as accepting Wilson's (controversial) unification thesis, but we will take on board the view that the SEM framework, including Wilson's suggestion to look for the presence of law-like background conditions, is a powerful tool to analyze the nature of dependence relations, especially where standard analyses are in trouble, such as in the case of frame-dragging in GR.

4 Frame-Dragging as a Nonstandard Determination Relation

We ended Sect. 2 by pointing out that counterfactual theories of causation are in trouble in the GR context. One can thus ask what progress we really made by adopting SEM's framework, given that it heavily relies on counterfactuals. In reply we note that, first of all, in this context the truth values of counterfactuals are not evaluated by comparing possible worlds, but by computing the values of the structural equations. In Sect. 2 we said that in GR it is not so straightforward to consistently and nonarbitrarily single out a model that represents a counterfactual situation to a given one (recall the disappearing masses case). This puts possible worlds semantics in trouble because it undermines the notion of "nearness" of possible worlds (which vacuum possible world is closest to ours?). In our case, we do not need any metaphysical notion of similarity among possible worlds to evaluate counterfactuals, we just need the posited structural relations among the relevant variables of the model. This shifts the arbitrariness involved in the evaluation of counterfactuals from a metaphysical to a methodological perspective, which is in fact closer to the way physicists engage in counterfactual reasoning.

The problems with interventionist counterfactuals are also mitigated because our counterfactual semantics is much more liberal than the standard interventionist one, allowing for any kind of physical and metaphysical intervention, possible or impossible. The worry that, say, a physically impossible intervention would render the analysis void because it would violate Einstein's field equations is unfounded, since the variables we intervene upon are part of the structural equations, whose adoption is of course justified by invoking the field equations, but which are nonetheless distinct from these field equations. In other words, by setting physically or even metaphysically impossible interventions on the variables, we are not feeding garbage into the theory's dynamics, thus getting garbage in return, but we are just testing the set of relations internal to our structural model. It might be the case that, in the end, the model turns out to be at odds with GR in one way or the other (e.g., local curvature dropping to zero when mass is increased), which only means we have to modify the structural model.

To make this point more vivid, let us consider first the problem of trans-world identification. Clearly, the use of this methodology would dramatically ease the evaluation of counterfactual changes. Indeed, now it is no more matter of looking how, say, the *very same* region S of spacetime in a possible world W looks in a near world W', because any counterfactual change of S translates into the corresponding change of values of the variable s representing S in the corresponding structural equations. Furthermore, let us consider counterfactuals involving small local changes in a model of GR that leave the rest of the universe untouched. In this case, all we have to do is to take an appropriate (e.g., fine-grained enough) structural model, "zoom in" the particular situation by looking at the variables that describe the local state of affairs, change their values as required, and then see if and how this change affects the "global" variables in the structural equations. Obviously, we cannot perform such a manipulation directly on a solution of Einstein's field equations. However, we can use the structural model to "transition" from the starting solution of the field equations (compatible with the starting values of the variables figuring in the structural equations) to a new one which encloses the changed state of affairs. The necessary condition is that the structural model be fully compatible with the laws of GR, otherwise the changed state of affairs fixed by the new values of the structural variables might not be encoded in any solution of the field equations. Note that this "transitioning" between solutions of GR is the SEM counterpart to Lewisian possible worlds' vicinity. Now, however, the similarity comparison between different states of affairs is not justified by metaphysical considerations, but just by pragmatic ones (i.e., that the structural model linking the two solutions works well in capturing the "backbone" of dependencies for both situations). It is easy to see how this mitigates also the issue of evaluating the counterfactual situation in which all the masses in the actual universe disappear: just construct an informative enough structural model that rightly captures the dependencies in the actual world, set the material variables in the dependence chain to zero, and see which vacuum model of GR is compatible with this new situation.

The above discussion is very schematic and is meant to hint at the fact that the advocated SEM framework fares better than standard frameworks, at least in cases where it is already clear what one is looking for (the frame-dragging case being one of those, as we will see in a moment). However, we are not claiming that the generalized SEM framework is the panacea for the troubles with counterfactual analyses of dependencies in GR. This line of research is still in its infancy, so a detailed assessment of how well the SEM framework fares *in general* in GR is still to come.

Having argued that the SEM framework is a promising approach to counterfactual reasoning in GR, we can now focus on the concrete case of rotational frame-dragging, and see if Wilson's analysis in terms of metaphysical/nomologic dependence helps us in judging whether causation is involved. For sure, by using Wilson's law-based characterization of dependence, the question of whether the dependence relation between the angular momentum $\mathbf{J}$ of the rotating mass and the angular velocity $\mathbf{\Omega}$ of the precessing local frame is instantaneous or retarded becomes irrelevant to the causal verdict: $\mathbf{\Omega}$ would causally depend on $\mathbf{J}$ just in case we have to invoke the laws of nature in order to justify the structural equation relating them. Put in these terms, the verdict seems to be trivial: yes, $\mathbf{\Omega}$ and $\mathbf{J}$ are causally related because in the analysis of the dependence relation involved we must invoke Einstein's field equations. However, this verdict would be too hasty. In fact, as we are going to see, the graph connecting these two variables has many more nodes to be taken into account.

In Sect. 2, we have seen that all cases of rotational frame-dragging follow the same scheme –although the degree of Machianity of each case depends on the particular physical setting considered. We start by giving information about the material distribution (the equation of state involving density ρ and pressure p of matter, which in turn determines the stress energy tensor $T_{\mu\nu}$). We then give the set of background symmetries (with related Killing vector fields) compatible with $T_{\mu\nu}$ (for example, external cylindrical symmetry for the Einstein–Lense–Thirring case). We then compute the angular momentum via (9), and from that we construct a linear perturbation $\delta h_{\mu\nu}$ of the background. We use the stress-energy tensor to calculate the perturbation via Einstein's field equations. In virtue of the symmetries of the background, we apply Noether's theorem to find that only the $0i$ $(i = 1, 2, 3)$ part of the field equations are involved. Thus, we find the form of the dragging potential f^a by (16) (or the particular cases (12) and (14)) and, from this, we construct the vector potential $\mathbf{A} = (0, 0, f^a)$ by fixing a convenient gauge (e.g., the Lorenz gauge). Finally, we get to the angular velocity of precession $\mathbf{\Omega}$ of a local inertial frame via (10), again by a choice of gauge that fixes the value of the function e^{α}. Hence, the graph of the structural model is:

$$\boxed{\rho, p} \xrightarrow{(i)} \boxed{T_{\mu\nu}} \xrightarrow{(ii)} \boxed{\mathbf{J}} \xrightarrow{(iii)} \boxed{f^a} \xrightarrow{(iv)} \boxed{\mathbf{A}} \xrightarrow{(v)} \boxed{\mathbf{\Omega}}$$

Of course, the actual structural equations relating these variables would be extremely complex. However, for our purposes, it is sufficient to evaluate the mediating principles invoked at each step. They are:

(i) State equation of matter (analytic functional dependence).
(ii) Equation (9) (analytic functional dependence) plus extra-theoretical (mathematical) symmetry considerations (e.g., Birkhoff theorem).
(iii) Einstein's field equations (law of nature), weak field approximation, Noether's theorem (mathematical law), Eq. (16) plus boundary conditions (e.g., $\omega(\infty) = 0$ or closedness).
(iv) Gauge fixing.
(v) Equation (10) (analytic functional dependence), gauge fixing.

We submit that, given the treatment of rotational frame-dragging presented in Sect. 2, the above model faithfully captures the relevant physical variables and their dependence for the phenomena in question. We also claim that the above model is much more powerful and useful for causal analysis purposes than standard counterfactual frameworks. In fact, our model can be used either to analyze the general type of dependence relations in the formal schema of rotational frame-dragging, which includes all cases from the Einstein–Lense–Thirring effect to general FLRW Machian cosmologies, or to analyze the token relations in each of the particular cases just mentioned. For comparison, Lewis' theory can only handle token cases. The great enhancement in analytical power is evident when we try to assess the Machianity of the different occurrences of frame-dragging. First of all, the Machian hallmark of frame-dragging is given by the fact that the exogenous variables represent material degrees of freedom, making manifest Mach's idea of material origin of inertial effects. Moreover, all the considerations we have dispersed here and there in Sect. 2 are beautifully summarized in the list of mediating principles (i)–(v). The less Machian cases feature "absolute" boundary conditions in (iii), while more Machian cases feature a condition of closedness in (iii), or even no boundary condition at all.

Coming back to the causality question, it is now clear why the standard counterfactual analyses of the dependence relation between $\mathbf{J}$ and $\mathbf{\Omega}$ performed so far gave an indecisive verdict—and why we cannot equate frame-dragging to everyday instances of causation, such as the stone/window case. Basically, the standard approaches systematically overlook both the fine-grained linking structure between $\mathbf{J}$ and $\mathbf{\Omega}$ and the dependence chain upstream of $\mathbf{J}$. So what is the verdict delivered by the SEM framework? We note that all links except for (iii) are not strictly speaking justified by laws of nature. The state equation for matter might count as a law in so far as it constrains ρ and p, but it just establishes a functional dependence between them and the stress-energy tensor. The abundance of nonlaw-like mediating principles hints at the fact that the dependence involved is much more logical/metaphysical than nomological. However we cannot speak of a pure case of grounding here because the middle link (iii) involves Einstein's field equations. (The nomological link here "corrupts" the whole chain, in much the way that a single stochastic link in an otherwise deterministic mechanism makes the whole mechanism's operation stochastic.) Furthermore, (iii) is the most important link in the chain, since it connects the upstream variables describing the overall material distribution with the downstream variables accounting for the (geometrical) properties of local inertial frames. Thus, we are in a situation where the "heart" of the chain depicts a nomological kind of dependence, while the rest of the links point at a metaphysical one. In our opinion, this mixed chain is the sign that a nonstandard determination relation is at work here.

The other reason to view the determination relation here as not a purely causal relation has to do with the invertible nature of the connection between inertial (metrical) structure and matter distribution in GR. Although the specific structure we laid out above is not directly invertible (that is, one can't input an $\mathbf{\Omega}$ and derive the state of matter ρ, p by working the mathematical derivations in reverse), it is nevertheless the case in GR that the metric field directly determines the material distribution via Einstein's equations. Therefore, one could presumably construct a closely related causal graph starting with a variable representation of the metric with more or less local precession and ending with a determinate quantity $\mathbf{J}$ of angular momentum in the appropriate coordinate gauge. Once again, the field equations, which are law-like, would play a key role in justifying one of the middle links of the graph, yet we would feel no temptation to consider the graph as establishing a causal connection between $\mathbf{\Omega}$ and $\mathbf{J}$.

The fact that the graph would link things that are simultaneous in the chosen coordinate frame or frames is not the reason why we would be reluctant; the explanation lies elsewhere, in deeply held physical intuitions about how spacetime structure and matter present in spacetime may relate to each other. This is not the place to explore those intuitions, and whether they ought to be defended or questioned. Our point is that the mathematical connections may be just as tight if we run them in the opposite direction, from a precessing metric to a quantity of angular momentum in the matter distribution about the central point. Despite this, an intervention on the central metric entailing a concomitant change in distant angular momentum does not feel like a connection that reveals causation, but rather at most a law-like codependence relation. If we judge by the similarity of the two dependence stories, we should judge them alike: both as amounting to a nonstandard determination relation that should not be held to be a clear case of causation.

We finish this section by considering a possible no-go objection against our reasoning. The objection goes like this: the frame-dragging effects we are discussing are just the result of an approximation, so we should not read too much into them, especially from a metaphysical perspective. To this objection, we reply that the approximation works extremely well for physically realistic situations, where small velocities and weak gravitational fields are involved, to the point that no higher-order corrections are needed to achieve empirically adequate results (see Everitt et al., 2011). This is, we think, more than a sufficient reason to investigate what kind of goings-on is captured by (16), especially in light of the fact that nobody, to our knowledge, has been so far able to come up with a higher-order derivation of frame-dragging effects in GR.

5 What Kind of Dependence?

We have spent much of the previous section establishing what the dependence relation underpinning frame-dragging effects is not. It is not straightforwardly causal, but it has too much physical import to be considered just metaphysical. Therefore, it seems that we are looking at a strange new animal lurking behind the bushes. The question then is: what is a frame-dragging relation? As we are going to see, the answer to this question is not univocal and is open to debate. Roughly speaking, we can isolate two possible responses.

The first response, which we might call *monist*, takes to its extreme consequences Wilson's unificatory thesis for dependence relations. Hence, this brand of monism would deny that the dependence depicted by frame-dragging is half-way between causation and grounding. This is because there is just *one* dependence relation—call it causation with a big "c", grounding with a big "g", or what have you—so the metaphysics of windows shattering, that of singletons, and that of frame-dragging is one and the same. The reason why, so far, metaphysicians had the impression of dealing with conceptually different cases is that they focused too much on details that cut no ontological ice. From this perspective, the SEM framework is particularly useful in pointing out what—according to monism—went wrong in the standard analytical approach. In a nutshell, the issue boils down to the mediating principles justifying a structural model. To see this, let's go back to the stone/window and member/singleton cases. Both circumstances are modeled in the very same way: the same variables, the same structural equations, and the same counterfactual pattern instantiated. In other words, the SEM framework simply does not distinguish between the two dependencies. The only demarcation is given by the mediating principles involved, but such principles are *external* to the analytical framework exploited. Therefore, in some sense, mediating principles are just a ladder used to reach the structural model and then kicked off once the model is complete. As such, these principles have a pragmatic rationale but very dubious metaphysical import. Another cue that might lead in this direction is that there is no consensus over the metaphysical status of laws of nature, some philosophers arguing that they are just another brand of metaphysical principles (e.g., necessitarianism), some others going as far as claiming that there is no such thing as laws of nature. Clearly, monism is an appealing thesis in light of this debate, since it dispenses with the need of providing an account of laws of nature in order to spell out what a dependence relation is. On the other hand, monism has a huge downside in that it might be accused of solving the issue by trivializing it. Indeed, dismissing the prima facie huge conceptual differences among dependence relations as descriptive fluff seems a rather unsatisfactory move. Usually, we clarify things up by adding details to the description, not by blurring the whole picture.

Hence, one might consider an opposite, *pluralist*, response to the original question. Pluralism differs from monism in that it takes the demarcation by mediating principles very seriously from a metaphysical point of view. Thus, under a pluralist reading of the SEM framework, causation and grounding are distinct but related concepts. This means that dependence relations are not sparse, but can be arranged in a scale depending on the type of principles mediating such dependencies. Hence, on one side sits causation, which is mediated by laws of nature only,[1] while on the opposite side sits grounding, which is not mediated at all by laws of nature. In this picture, the claim that frame-dragging dependence lies between causation and grounding is literally true, since it is only partially mediated by laws of nature. In this sense, frame-dragging dependence might be (unimaginatively) dubbed "caunding" or something like that. The problem with this line of reasoning is that it does not clarify whether such a dependence scale has just three positions (nomological, mixed, and metaphysical) or it covers a wider spectrum capturing the degree of mixing. In this second case, a further criterion should be provided for "counting" or "weighing" the law-like links in a dependence chain, in order decide the position in the spectrum of the corresponding relation. Lacking this additional criterion, the

[1]Some people may be skeptical that, e.g., cases of causation by omission fulfill this criterion. However, for simplicity's sake, we will gloss over this contentious issue.

question regarding the nature of frame-dragging dependence remains only partially answered. Contrary to monism, pluralism seeks indeed to provide a nontrivial, informative account of dependencies by adding more details to the picture. However, the added layer of complexity seems to bring new issues to the fore, rather than settling the already existing ones. From this point of view, it is contentious whether pluralism fares really better than monism.

6 Conclusion

Standard analyses of dependence relations are often aimed at well-behaved, clear-cut cases taken from everyday life. For example, all of us have a clear grasp of what it means for a lightning to determine a fire in the woods, or for a marble to be colored in virtue of being red. These analyses commonly exploit tools such as Lewisian counterfactuals, and rely on demarcation criteria given in terms of distinctions like events/facts, temporality/fundamentality, synchronicity/diachronicity, and contingency/necessity. In this paper, we have tried to show that when it comes to fundamental physics, and GR in particular, things are not so straightforward. The lesson we have drawn from this fact is that, perhaps, we should follow a new path in metaphysical theorizing. The blazing of such a trail has recently started with the work of Schaffer, Wilson, and others on a generalized SEM framework that unifies the—so far, disjoint—analyses of causation and grounding. We have applied this framework to frame-dragging dependencies in GR, showing that it is possible to deliver a (partial) characterization of the underlying dependence relation, which is at least conceptually clearer than that delivered by standard analytical frameworks. Of course, much more has still to be said regarding such a relation in order to reach a full metaphysical characterization. However, as discussed in the previous section, this is part of the broader issue of making full metaphysical sense of the tools provided by the generalized SEM framework (in particular of the demarcation criterion in terms of the laws of nature/metaphysical principles distinction). As such, our work is just another preliminary step toward a full understanding of this nonstandard framework. A further step would be to investigate the nature of the dependence relation underpinning geodesic motions in GR, looking for possible differences with the frame-dragging case. In the end, the hope is that a fully developed analysis of dependencies in spacetime physics might help in solving the thorny issue of providing a clear metaphysical story for the emergence of classical spacetime from an underlying quantum-gravitational regime. At this stage, however, fulfilling this latter task lies far in the future.

Acknowledgments We are grateful to an anonymous referee and the editors of this volume for their comments on an earlier draft of this paper. One of us (AV) worked on this paper while being at the University of Barcelona as a Juan de la Cierva fellow. Hence, the financial support from the Spanish Ministry of Science, Innovation and Universities, fellowship IJCI-2015-23321, is gratefully acknowledged. Research for this paper was also supported by the Spanish MEIC research project FFI2016-76799-P.

References

Barbour, J. and H. Pfister (Eds.) (1995). *Mach's principle: from Newton's bucket to quantum gravity*. Birkhäuser.

Bičák, J., J. Katz, and D. Lynden-Bell (2007). Cosmological perturbation theory, instantaneous gauges and local inertial frames. *Physical Review D 76*, 063501.

Brill, D. and J. Cohen (1966). Rotating masses and their effect on inertial frames. *Physical Review 143*(4), 1011–1015.

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

Curiel, E. (2015). If metrical structure were not dynamical, counterfactuals in general relativity would be easy. https://arxiv.org/abs/1509.03866.

Everitt, C., D. B. DeBra, B. Parkinson, J. Turneaure, J. Conklin, M. Heifetz, G. Keiser, A. Silbergleit, T. Holmes, J. Kolodziejczak, M. Al-Meshari, J. Mester, J. Mester, B. Muhlfelder, V. Solomonik, K. Stahl, P. Worden Jr., W. Bencze, S. Buchman, B. Clarke, A. Al-Jadaan, H. Al-Jibreen, J. Li, J. Lipa, J. Lockhart, B. Al-Suwaidan, M. Taber, and S. Wang (2011). Gravity Probe B: Final results of a space experiment to test general relativity. *Physical Review Letters 106*, 221101.

Hoefer, C. (2014). Mach's principle as action-at-a-distance in GR: The causality question. *Studies in History and Philosophy of Modern Physics 48*, 128–136.

Huggett, N. and C. Hoefer (2017). Absolute and relational theories of space and motion. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2017/entries/spacetime-theories/.

Lam, V. (2010). Metaphysics of causation and physics of general relativity. *Humana.Mente 13*, 61–80.

Lindblom, L. and D. Brill (1974). Inertial effects in the gravitational collapse of a rotating shell. *Physical Review D 10*(10), 3151–3155.

Lynden-Bell, D., J. Katz, and J. Bičák (1995). Mach's principle from the relativistic constraint equations. *Monthly Notices of the Royal Astronomical Society 272*, 150–160.

Mach, E. (1883). *Die Mechanik in ihrer Entwickelung, historisch-kritisch dargestellt*. Brockhaus. English translation by T.J. McCormack. Open Court Publishing Co., 1893.

Misner, C., K. Thorne, and J. Wheeler (1973). *Gravitation*. W.H. Freeman and Company.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pfister, H. (2007). On the history of the so-called Lense-Thirring effect. *General Relativity and Gravitation 39*(11), 1735–1748.

Pfister, H. and K. Braun (1985). Induction of correct centrifugal force in a rotating mass shell. *Classical and quantum gravity 2*(6), 909.

Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies 173*, 49–100.

Wilson, A. (2017). Metaphysical causation. *Noûs*. https://doi.org/10.1111/nous.12190.

Wilson, A. (2018). Grounding entails counterpossible non-triviality. *Philosophy and Phenomenological Research 96*(3), 716–728.

Woodward, J. (2016). Causation and manipulability. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2016/entries/causation-mani/.

Approximate Local Poincaré Spacetime Symmetry in General Relativity

Samuel C. Fletcher

Abstract How does general relativity reduce, or explain the success of, special relativity? Answering this question, which Einstein took as a desideratum in the formulation of the former, is of acknowledged importance, yet there continues to be disagreement about how exactly it is best answered. I advocate here that part of the best answer involves showing that every relativistic spacetime has an approximate local Poincaré spacetime symmetry group, the spacetime symmetry group of Minkowski spacetime. This explains the application of Minkowski spacetime concepts that depend on, e.g., the conserved quantities that spacetime symmetries guarantee. I contrast this approach with another that instead invokes the strong equivalence principle, which focuses on the distinct notion of Poincaré invariance of dynamical equations. After showing with some examples that neither notion is necessary for the other, I use those examples to illuminate contrasting positions on the explanatory role of local approximate spacetime geometry, defending Torretti (1996) against criticisms by Brown and Pooley (2001). Finally, I acknowledge that establishing approximate local Poincaré spacetime symmetry is not a complete answer to the explanatory question with which I led, discussing in the concluding section further work that could lead to a complete answer. This includes specifying the circumstances under which matter fields in a general relativistic spacetime "behave" locally like those in Minkowski spacetime.

1 Introduction: Explanation, Reduction, and Symmetry

1.1 Explanation and Reduction

Several constraints and heuristics guided Einstein's endeavor to find an acceptable relativistic theory of gravitation. One was such a theory's relationship to the special theory of relativity (SR). Not only should it reduce to SR when substantial

S. C. Fletcher (✉)
University of Minnesota, Minneapolis, MN, USA
e-mail: scfletch@umn.edu

C. Beisbart et al. (eds.), *Thinking About Space and Time*, Einstein Studies 15,
https://doi.org/10.1007/978-3-030-47782-0_12

gravitational fields were absent (Renn and Sauer, 1998, p. 97), but also the interpretation of its metric components should mirror that for SR (Norton, 1984, p. 261), and it should reproduce the kinematics of relativistic continua (Janssen, 2014, p. 211). In a word—though, perhaps, not one Einstein would have used—a relativistic theory of gravitation had to *explain* the success of SR. Despite this potential anachronism, it captures succinctly the demand incumbent on a relativistic theory of gravitation: to answer how applications of SR were as successful as they indeed were, even though they idealized away all gravitational effects acknowledged as ever present.

There is no debate about whether the culmination of Einstein's endeavor in the general theory of relativity (GR) fulfills this demand. It does, but *how* it best does it not often agreed upon. My goal in this note is to exhibit in some novel detail *one* aspect of my preferred explanation, which is to delineate the circumstances under which successfully applied (partial) solutions, or models, of fields in SR approximate those in GR that are in those circumstances at least as successful. In this sense, GR *reduces to* SR. It is not a reduction in the usual philosophers' sense of deducing SR from GR, for the two are generally incompatible when applied to the self-same phenomena. Rather, it is somewhat more in the vein of "physicists' sense of reduction" as described by Nickles (1973) and elaborated by Ehlers (1986): it shows how the (typically) newer, more expansive theory represents more phenomena successfully, and accounts for how the (typically) older, less expansive theory represented phenomena successfully by showing that it well approximates the former in these cases.

All that said, this is not the occasion for a general disquisition on the concepts of reduction or explanation. I shall not defend the claims that accounting for the success of SR within GR is a form of both, but I also shall not draw from accounts of explanation or reduction in my arguments. So for those skeptical of these claims, my use of these terms may serve as a label for the elaboration of how the domain of application of SR is subsumed through approximation and correction to that of GR.[1]

1.2 Spacetime Symmetries

Another reason I do not focus on defending these claims is that I only treat presently a fragment of the whole project of reducing or explaining the success of SR. In particular, I focus on the successful application of *spacetime symmetries*. Given any

[1]Rosaler (2019) declines from calling this sort of reduction-as-domain-subsumption an explanation, citing the deep controversies over accounts of explanation. I agree with him that the present sense of explanation is not well modeled, e.g., within the deductive-nomological or other standard philosophical accounts of scientific explanation, but so much the worse for those accounts.

relativistic spacetime—that is, a model (M, g), with M a four-dimensional smooth manifold and g a smooth Lorentz signature metric on it[2] —consider the collection of diffeomorphisms $\psi : M \to M$ such that the pushforward of the spacetime metric is itself, i.e., $\psi_*(g) = g$. In a word, a spacetime symmetry is an automorphism of a relativistic spacetime; it maps the spacetime metric back onto itself within the same spacetime.[3] The collection of all such maps forms a group under composition, hence constitutes the *spacetime symmetry group* of (M, g). There are many fewer spacetimes with symmetries than those without, so their presence greatly restricts the range of possible states of affairs and spacetime structures.

When a one-parameter family of spacetime symmetries is generated infinitesimally from a smooth vector field κ on M, it also gives rise to conserved quantities.[4] In more detail, suppose that κ Lie-derives the metric g, i.e., $\pounds_\kappa g = 0$, in which case κ is known as a Killing vector field.[5] Killing vector fields can give rise to two sorts of conserved quantities, one associated with the worldlines of free particles, and another associated with divergence-free, localized symmetric tensor fields, like those representing energy-momentum (Malament, 2007, §2.7).

Consider first any particle with worldline $\gamma : I \to M$, where $I \subseteq \mathbb{R}$, and tangent vector field ξ on $\gamma[I]$. If γ is a geodesic and κ is a Killing vector field, then $J = \xi^a \kappa_a$ is constant on $\gamma[I]$. These conserved quantities can often be interpreted as energy, linear, and angular momentum, and so on. Consider second any symmetric tensor field T^{ab} that is divergence-free, i.e., $\nabla_a T^{ab} = 0$. Suppose also that it is localized, in the sense that it vanishes outside some timelike world tube (Malament, 2007, p. 255). Then one can integrate $T^{ab}\kappa_a$ over any spacelike slice of the tube, such that the resulting quantity is constant across slices and independent of slicing. These quantities can often be interpreted analogously to those for point particles. This provides a sense in which localized but extended bodies also have conserved quantities as determined by these spacetime symmetries.

A general relativistic spacetime may not have any non-trivial such spacetime symmetries (i.e., its spacetime symmetry group may consist only of the identity). By

[2] I also assume that M is connected, Hausdorff, and paracompact, and that the metric signature is $(+---)$. Throughout, roman sub- and superscripts denote abstract indices, while Greek and numerical ones denote components in a contextually specified basis. (See, e.g., Wald (1984, §2.4) for more on abstract index notation.) When an expression does not involve index contraction, I will often omit the indices to reduce notational clutter when no confusion should arise from doing so.

[3] I am implicitly using the identity map on M to compare the image of the pushforward with its argument. This is ultimately a convention: there is nothing mathematically or representationally privileged about the identity over any other diffeomorphism of the manifold, but choosing a different standard of comparison would yield an entirely representationally equivalent set of spacetime symmetries (Fletcher, 2020). This is all because diffeomorphisms are the isomorphisms in the category of smooth manifolds.

[4] Not all spacetime symmetries are such: consider so-called discrete symmetries such as reflections or time-reversal.

[5] This condition is equivalent with κ satisfying Killing's equation, $\nabla_{(a}\kappa_{b)} = 0$, where ∇ is the Levi-Civita covariant derivative operator compatible with the metric g. The Killing vector fields also form a Lie algebra, which will play a role in Sect. 3.

contrast, Minkowski spacetime has ten linearly independent Killing vector fields, the maximal number that a spacetime can have (Hall, 2004, Theorem 10.2.iii). These fields generate the symmetry group known as the *Poincaré group* (or, less elegantly, the inhomogeneous Lorentz group). It consists of four one-parameter families of symmetries corresponding to (spatial and temporal) translations, three corresponding to spatial rotations about a fixed axis, and three corresponding to boosts, which can be understood as linear transformations of Minkowski spacetime considered as an affine space. These symmetries are important in SR because their presence corresponds with both a restriction on which types of matter dynamics are possible and the presence of conserved quantities. The standard coordinates of Minkowski spacetime in SR are well-adapted to these symmetries: the congruences of the coordinate fields are integral curves of the spacetime symmetries, and the metric components expressed in their values are independent of them (Hall, 2004, p. 294–5).

Clearly many successful applications of SR appeal to these symmetries and conserved quantities, yet when de-idealized and treated within GR, these symmetries disappear. How does one thus account for the successful but idealized application of special relativity in these circumstances? I draw on a novel account of approximate spacetime symmetries (Fletcher, 2018) to show, in Sect. 3.1, that every relativistic spacetime has approximate local Poincaré symmetry. Although a generic spacetime of GR has no non-trivial spacetime symmetries, there is nevertheless a sense in which every spacetime approximately has maximal spacetime symmetry, if only locally. Readers thoroughly familiar with relativity theory ought not be surprised by this conclusion, but what is novel in my account is the specific details of what this means and how it comes about.

1.3 Outline of the Remainder

To motivate my account, I first consider another approach to understanding the "local validity" of SR in Sect. 2, one based on the so-called strong equivalence principle. Such approaches are venerable in the foundational literature on spacetime theories, so one might wonder why the new formal apparatus I seek to introduce is really necessary. I show that while, superficially, this approach via the strong equivalence principle is concerned with the same question, on closer examination it is concerned with one that is slightly different, about Poincaré *invariance*—namely, what sorts of coordinate transformation preserve the form of certain equations. After presenting in Sect. 3.1 how one can account in part for the success of SR in terms of approximate local spacetime symmetries, I describe in Sect. 3.2 how Poincaré invariance of equations is neither necessary nor sufficient to account for the successful application of Poincaré symmetry in local regions of a generic relativistic spacetime.

Due to the subtly different explanandum to which the strong equivalence principle is typically applied, perhaps this should not come as a great surprise.

But, in the slightly digressive next section (Sect. 4), I show how appreciating these examples sheds light on an old debate between Torretti (1996) and Brown and Pooley (2001) about what local approximate *geometry* does and does not explain. Torretti claims that the local approximate Minkowski geometry of any relativistic spacetime entails the local Lorentz invariance of dynamical laws "referred to local Lorentz charts," while Brown and Pooley deny this. This new light vindicates Torretti, properly interpreted, from the charge of a *non sequitur*.

Although the application of Poincaré symmetry in local regions of a generic relativistic spacetime accounts in part for the success of SR, it does not account for it in total. Thus, in the concluding Sect. 5, after summarizing the preceding sections, I discuss some of the limitations of focusing on spacetime symmetries only. In particular, it does not necessarily account for the well-approximation of the values of observable matter fields in a generic GR spacetime by those of corresponding fields in (a region of) Minkowski spacetime. This naturally suggests extensions of the present strategy to these fields, which I outline, leaving the details, however, for future work.

2 The Strong Equivalence Principle

There is a venerable tradition of considering the relationship between the special and general theories of relativity in terms of an equivalence principle (Pauli, 1958). Although the equivalence principle is rightly attributed to Einstein, he took the principle rather to be a covariance principle applied in Minkowski spacetime only, which allowed one to interchange uniformly accelerated frames with unaccelerated ones in a uniform gravitational field (Norton, 1993, §4.1); he did not see it as linking the special and general theories (Einstein, 1923, 1956). Advocates of such a link have acknowledged this (Read et al., 2018, p. 14n2), emphasizing that one should understand the relevant version as the *strong* equivalence principle (SEP), rather than *Einstein's* equivalence principle. Here is a typical statement by Brown (2005, p. 169):

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

This raises at least three interpretive questions:

1. What, exactly, are laws governing fundamental non-gravitational interactions?
2. What does it mean that "tidal gravitational effects can be ignored"?
3. What is does it mean for a law to take "special relativistic form," or its "simplest form"?

Read et al. (2018, p. 17) clarify further each of these in their restatement of the SEP: "The dynamical equations for non-gravitational fields reduce to a Poincaré invariant form, with no terms featuring the Riemann tensor or its contractions, in a neighborhood of any $p \in M$." Thus they provide the following answers:

1. The laws governing fundamental non-gravitational interactions are the dynamical equations for matter fields, those fields with an associated energy-momentum tensor (Read et al., 2018, p. 14n1).
2. Tidal gravitational effects are terms in the aforementioned dynamical equations that make explicit reference to curvature; justifiably "ignoring" them requires sufficiently low sensitivity of one's experimental apparatus to curvature effects that they cannot differentiate between the presence and absence of curvature (Read et al., 2018, p. 17); thus this clause can only be satisfied contextually (Brown, 2005, p. 170).
3. An equation takes special relativistic form when it is invariant under Poincaré transformations; it takes its simplest form when the number of terms in the equation does not reduce by being expressed in a special local inertial frame (Read et al., 2018, p. 21).

I will return to further elucidation of the third answer shortly; regarding the second answer, Brown (1997, p. 72) emphasizes the context also includes the size of the region under consideration—a region $U_\epsilon(p)$ containing p where one can find the aforementioned locally inertial coordinates. These are coordinates in which the connection components within $U_\epsilon(p)$ are sufficiently small that, for geodesic worldlines of free point particles intersecting this region, their coordinate expressions do not measurably deviate from straightness. (In other words, their coordinate velocities are approximately constant.) So, when Brown (1997, p. 71) writes that "In GR, for all regions of spacetime in which curvature can be ignored, SR is valid by fiat," the decree to which he alludes is the SEP as stated above.

Does Brown's claim about the local validity of SR include spacetime symmetries? Some of his comments suggest that it should:

> It should be noted also that with respect to $U_\epsilon(p)$ it is perfectly legitimate to define "local" symmetry groups that contain as a sub-group the set of spatial translations, such as the Poincaré group in SR. This calls into question the frequent claim that it is the Lorentz group, rather than the Poincaré group, which represents the local symmetries in GR. (Brown, 1997, p. 79n17)

However, here we must return to the third answer above to adjudicate whether the local validity of SR, as the SEP purportedly grants, guarantees the existence of local Poincaré spacetime symmetries. As Read et al. (2018, p. 19) explain, a problem arises with their interpretation, if they were to be taken as spacetime symmetries properly understood: on the one hand, for any event p SR is supposed only to be valid in some $U_\epsilon(p)$, yet the Poincaré group contains actions (such as translations) that seem like they ought to map p to a point q outside of $U_\epsilon(p)$; thus understood, some spacetime symmetries are not symmetries applicable to $U_\epsilon(p)$ at all. Put plainly, the symmetries of $U_\epsilon(p)$ cannot be the Poincaré symmetries because that group is not even well-defined on a bounded region of spacetime. "The resolution is to view the Poincaré transformations discussed above as passive only—their action is not on spacetime points at all, but rather on the chart space, i.e., the codomain of the coordinate charts" (Read et al., 2018, p. 19n32). In other words, the normal (and "locally inertial") coordinate system assigned to $U_\epsilon(p)$ to which the SEP refers is a fragment of a Lorentz chart, and the Poincaré transformations invoked act on these charts, shifting the coordinates assigned to points of $U_\epsilon(p)$ to one related by the invoked Poincaré transformation. No diffeomorphisms or pushforwards are invoked that would map points of $U_\epsilon(p)$ outside it, but this is not needed to account for the invariance of form of equations referring only to points of $U_\epsilon(p)$.

This makes the invocation of the SEP for its stated aims salutary and consistent.[6] But it does not explain or account for local spacetime symmetries. Those do not concern directly the preservation of forms of equations, which is a mostly syntactic notion, but rather the invariance of spacetime structure, which is a mostly semantic notion. Moreover, as I show in Sect. 3.2, the SEP holding approximately on a spacetime region, to any given degree of approximation, is neither necessary nor sufficient for that region to have even approximate local Poincaré spacetime symmetry.

One possible reply to this conundrum would be to re-define what it means for a theory to be "special relativistic." Read et al. (2018, §7) describe, but do not advance, one way to do this: take such theories to be characterized solely by the constraint that "The dynamical laws governing matter fields are Poincaré invariant," instead of also assuming that "The inertial frames are global" (Read et al., 2018, p. 22), i.e., that the covariant connection is flat (and arises from a geodesically complete metric on $\mathbb{R}^4$ (Read et al., 2018, p. 22n45)). The advantage they cite for this position is that it allows one to claim that SR is locally valid exactly, instead of merely approximately.

But this is a doubly Pyrrhic victory. By moving the goalposts so much closer, it elides much of why the local validity of SR was important in the first place. The point was never to find an interpretation of the sentence "SR is locally valid" that makes it true, but rather to explain why SR is as successful as it is, despite its idealizations, on its own terms. SR as such employs Minkowski spacetime, with full Poincaré symmetry, and the conservation laws that this entails. Re-construing SR does not explain the successful albeit approximate application of these symmetries to local regions. Moreover, it conceals the important insight that the extent of SR's

[6] That said, it does bring out a lacuna in the argument for the chronogeometric significance of the metric—the argument for why, according to the dynamical perspective on relativity theory (Brown, 1997; Brown and Pooley, 2001; Brown, 2005), the spacetime metric (perhaps only approximately) measures or surveys times and distances. The argument hinges on observing that "The symmetries of the dynamical laws governing non-gravitational fields in the appropriate local neighborhood ... coincide with the symmetries of the dynamical metric field in this neighborhood" (Read et al., 2018, p. 19), with the latter understood as spacetime symmetries in the sense I have discussed (Read et al., 2018, p. 19n25). While the symmetry groups coincide, they act on different objects: the former acts on coordinates assignments to points and fields in a fixed region, while the latter acts on spacetime points and fields thereon. Why should this coincidence of two different types of objects deliver the interpretation of one?

success is based on approximation, not exact correspondence, and is limited in general to a neighborhood of an event. Without this insight, it becomes entirely obscure why and when SR is in certain circumstances empirically *in*adequate in comparison to GR.

3 Poincaré Symmetry

In the previous section, I invoked general features of approximate spacetime symmetries qua spacetime symmetries in my argument that they are not adequately captured by the SEP. But I had not yet given a precise definition of what it means for a symmetry to be local, or—more importantly—approximate. That is the task of the Sect. 3.1. Then Sect. 3.2 will introduce examples showing that Poincaré invariance of equations, as invoked by the SEP, is neither necessary nor sufficient for Poincaré spacetime symmetry.

3.1 Local and Approximate Spacetime Symmetries

In Sect. 1.2, I introduced a spacetime symmetry of a spacetime (M, g) as an element of a collection of diffeomorphisms ψ on M such that $\psi_*(g) = g$, and which form a group under composition. While this definition is slightly more general than is usually found in textbook treatments, which typically focus on symmetries that are generated as the flows of Killing vector fields (Malament, 2007, §2.7) or fields that satisfy generalizations of Killing's equation (Wald, 1984, Ch. C.3), it is also not general enough for present concerns in two ways. First, as the action of ψ is on all of M, it is a *global* symmetry, while an account of local symmetries on just proper parts of M is needed. Second, the condition that the action of the diffeomorphism preserves the metric, $\psi_*(g) = g$, is exact, while an account of approximate preservation is needed. These will be taken up and then combined in the next two subsections.

3.1.1 Local Spacetime Symmetries

The modification of the definition of global spacetime symmetries to yield that of local spacetime symmetries is quite simple. Let a spacetime (M, g) be given, and U, V be open submanifolds of M. The smooth map $\psi : U \to V$ is said to be a *local diffeomorphism* when it is a diffeomorphism of U considered as a manifold in its own right. If further $\psi_*(g) = g$, i.e., the pushforward of the metric on U along ψ yields the metric on V, then ψ is a *local spacetime symmetry* of (M, g) (Hall, 2004, p. 285). When $U = M$, ψ is also a global spacetime symmetry, but this is clearly a special case. For instance, consider any global spacetime symmetry of Minkowski spacetime, e.g., a time translation; when restricted to some open bounded (i.e., precompact) region, it is a local spacetime symmetry. If one transforms the Minkowski metric by, say, a conformal factor that is not constant outside of the domain and range of this local spacetime symmetry, it still remains a local spacetime symmetry, even though after making this transformation there are no non-trivial global spacetime symmetries.

The collection of all local spacetime symmetries for (M, g) does not typically form a group under composition because these symmetries in general do not share domains and codomains. However, whenever $\psi : U \to V$ and $\psi' : U' \to V'$ are local spacetime symmetries and $V \subseteq U'$, then $\psi' \circ \psi$ is also a local spacetime symmetry. In other words, the collection is closed under composition when the composition is well-defined. Moreover, the identity map on M counts as a local diffeomorphism, and because each local spacetime symmetry is a diffeomorphism of its domain, its inverse is a local spacetime symmetry as well. This yields a slight generalization of the group concept known as a *groupoid*. Thus, we may speak of the groupoid of all local spacetime symmetries for (M, g).

Local spacetime symmetries can be generated infinitesimally from smooth vector fields, just as global spacetime symmetries can, and give rise to conserved quantities in much the same way. The smooth vector field generating such a local symmetry for a region is called a local Killing vector field for that region. It must be defined at least on a larger region connecting the domain of the putative symmetry with its range. One important difference between local and global symmetries, however, is that the parameter for the flows that these local Killing vector fields generate may only be defined for a proper subinterval of $\mathbb{R}$. For example, in a spacetime describing gravitational collapse into a Schwarzschild-like black hole (Wald, 1984, p. 155–7), once the collapse has finished there is a local timelike (and hypersurface orthogonal) Killing vector field, i.e., the spacetime is locally static (Malament, 2007, p. 253). But that Killing vector field does not extend into the past, i.e., to the portion of spacetime during and before the collapse process.

Given a collection of local Killing vector fields for a region, one can classify what sort of spacetime symmetry group they model by examining the Lie algebra of the fields as they are defined on that region. Recall that the Lie bracket for the Lie algebra of such fields is defined through the Lie derivative, i.e., for vector fields α, β, γ defined on a common region U, their bracket is defined as $[\alpha, \beta] = \pounds_\alpha \beta$, which is anti-symmetric ($[\alpha, \beta] = -[\beta, \alpha]$) and satisfies the Jacobi identity,

$$[[\alpha, \beta], \gamma] + [[\gamma, \alpha], \beta] + [[\beta, \gamma], \alpha] = 0. \tag{1}$$

Moreover, the Lie bracket of local Killing vector fields on a region is a local Killing vector field on the same region. So, a region may be said in particular to have local

Poincaré spacetime symmetry when there is a collection of local Killing vector fields on that region whose Lie algebra is the Poincaré algebra.[7]

3.1.2 Approximate Spacetime Symmetries

An *approximate* (local) spacetime symmetry ψ should be a (local) diffeomorphism that satisfies the equation $\psi_*(g) = g$ approximately. To make sense of this demand, one must describe how a spacetime metric—or even a tensor field, more generally—approximates or is similar to another on a spacetime region. There is good reason to believe that there is no canonical way to do this, so we must determine from the present context of investigation what the relevant notion of similarity is (Fletcher, 2016). Although local symmetries may be defined on arbitrary submanifolds, I focus here on bounded (i.e., precompact) regions only. This restriction fits with the present goals because the ultimate explanandum here is the successful application of the spacetime symmetries of SR, which has been only in bounded regions (but much more than isolated points) of spacetime.

In this case, one can adapt the apparatus of the so-called compact-open topologies on Lorentz metrics, as developed in Fletcher (2014). To begin with, consider a precompact region U on which one is adjudicating the status of a putative collection of local spacetime symmetries. Next, consider a smooth *frame field* $\{t^a, x^a, y^a, z^a\}$ defined at least on the closure of the union of the images of U under those symmetry maps, recalling that a frame field is an collection of orthonormal vector fields, one of which (t^a) is timelike and the rest spacelike, that form a basis for the tangent space at any point. Such a frame field represents idealized temporal and spatial measuring instruments, and each of its component's local integral curves is a curve of constant temporal or spatial coordinate value. From it, one can construct the smooth (inverse) Riemannian metric[8]

$$h^{ab} = t^a t^b + x^a x^b + y^a y^b + z^a z^b, \tag{2}$$

which in turn defines a norm—what I will call the h-fiber norm—for covariant tensor fields, such as Lorentz metrics g_{ab}, at points where it is defined:

$$|g|_h = |h^{ab} h^{cd} g_{ac} g_{bd}|^{1/2}. \tag{3}$$

[7] I am eliding some inconsequential technicalities regarding the relationships between local diffeomorphisms, local transformation groups (associated with a connected Lie group), and infinitesimal transformation groups (associated with a Lie algebra). For more on these, including references, see Hall (2004, Ch. 5.11).

[8] In this mode of presentation, I have assumed that the region in question is temporally and spatially orientable, for this is equivalent to the existence of a frame field. (See footnote 11 for definitions of these properties.) However, even if the region did not have those properties, one can always start with some smooth (inverse) Riemannian that can be decomposed locally into a frame field.

Inspection shows that the h-fiber norm of g at a point returns the Frobenius norm of g expressed as a matrix in components of the originally chosen frame field, i.e., the square root of the sum of squares of these components. It serves as an aggregate measure of the magnitude of a field at a point as measured using the frame field. Furthermore, the definition of the h-fiber norm can be extended to fields with arbitrary covariant indices, $f_{a_1 a_2 \cdots a_n}$, as follows:

$$|f|_h = |h^{a_1 b_1} h^{a_2 b_2} \cdots h^{a_n b_n} f_{a_1 a_2 \cdots a_n} f_{b_1 b_2 \cdots b_n}|^{1/2}. \tag{4}$$

The significance of this extension is that it allows us to define a distance function between two metrics on a region U that compares the maximal differences not only of their components as expressed in the basis of the frame field, but also their derivatives up to order k:

$$d_U(g, g'; h, k) = \max_{j \in \{0, \ldots, k\}} \sup_U |\nabla^{(j)}(g - g')|_h, \tag{5}$$

where $\nabla^{(j)}(g - g')$ abbreviates $(g_{ab} - g'_{ab})$ for $j = 0$ and $\nabla_{c_1} \cdots \nabla_{c_j}(g_{ab} - g'_{ab})$ otherwise, with ∇ the Levi-Civita connection compatible with h.

With this apparatus in place, one can define an approximate (local) spacetime symmetry ψ on U. More precisely, a local diffeomorphism $\psi : U \to V$ (of course with $U, V \subseteq M$) is an *(h, ϵ)-spacetime symmetry to order k on U* when $d_U(g, \psi^*(g); h, k) < \epsilon$. (Note that when ψ is a member of a one-parameter family of local diffeomorphisms generated by a local Killing vector field κ, this is equivalent to the condition that $\sup_U |\pounds_\kappa \nabla^{(j)} g|_h < \epsilon$.) In other words, the maximum difference between the metric on U and the metric on the image of ψ, including its derivatives to order k, is no more than ϵ according to the frame field constructing h. More generally, one may say that U has an (h, ϵ)-spacetime symmetry groupoid G to order k on U when $\sup_{\psi \in G} d_U(g, \psi^*(g); h, k) < \epsilon$ and the elements of G form a groupoid. Further, when they are members of a one-parameter family of local diffeomorphisms generated by local Killing vector fields, one can classify this groupoid by the Lie algebra of the vector fields formed under the Lie bracket.

Theorem 1 *For every h, ϵ, and finite k as above, every point of every spacetime has a neighborhood on which there is (h, ϵ)-Poincaré spacetime symmetry to order k, where the usual group of symmetries is restricted to a groupoid that forms a neighborhood of the identity in the Poincaré Lie group.*

Proof Consider any spacetime (M, g), any event $p \in V \subseteq M$, and a diffeomorphism $\phi : U \to V$, where $U \subseteq \mathbb{R}^4$ is a neighborhood of Minkowski spacetime $(\mathbb{R}^4, \eta)$ such that $\phi_*(\eta)_{|p} = g_{|p}$. (It is always possible to satisfy this last condition because any tangent space at any point of each spacetime, equipped with a Lorentz metric, is isomorphic as an inner product space to any other such.) Consider also any Riemannian h defined at least on U, any $\epsilon > 0$, and any finite non-negative k. One can push forward onto V a collection of local Killing vector fields on U that form a basis for the Poincaré algebra. Each of these generates a one-parameter family of local flows ψ_t with flow parameter t. Since g is smooth, one can always find $V' \subseteq V$ that is a neighborhood of p and some $t' > 0$ such that for all $|t| < t'$ $d_{V'}(g, \psi_t^*(g); h, k) < \epsilon/10$. Since the Poincaré algebra is ten-dimensional, any linear combination of these local Killing vector fields with coefficients no greater than one will then generate a local flow $\tilde{\psi}_t$ such that $d_{V'}(g, \tilde{\psi}_t{}^*(g); h, k) < \epsilon$ for sufficiently small t, using the triangle inequality and the linearity properties of the Lie derivative. These linear combinations in turn, through the Lie exponential map, generate elements of the Poincaré Lie group in its identity component, which form a groupoid. Finally, because $\phi_*(\eta)_{|p} = g_{|p}$, these elements can be interpreted as translations, rotations, boosts, and combinations thereof, according to the corresponding interpretations of the elements of the Poincaré Lie group for Minkowski spacetime.

It remains to show that the elements of this groupoid satisfy the commutation relations of the Poincaré algebra on V'. For this, it suffices to prove that, if α and β are smooth vector fields on U, then $\phi_*([\alpha, \beta]) = [\phi_*(\alpha), \phi_*(\beta)]$. For in this case, it follows from any commutation relation $[\alpha, \beta] = \gamma$ for smooth vector fields on U that $\phi_*(\gamma) = \phi_*([\alpha, \beta]) = [\phi_*(\alpha), \phi_*(\beta)]$ on V, hence on V'.

The proof requires the application of three facts: for any smooth vector fields ξ, ζ and smooth scalar field f on U,

$$[\xi, \zeta](f) = \xi(\zeta(f)) - \zeta(\xi(f)), \tag{6}$$

$$\phi_*(\xi)(f)_{|p} = \xi(f \circ \phi)_{|\phi^{-1}(p)}, \tag{7}$$

$$\phi_*(\xi)(f) \circ \phi = \xi(f \circ \phi). \tag{8}$$

Equation 6 expresses the action, on a smooth scalar field, of the Lie bracket of smooth vector fields in terms of the commutator of directional derivatives of that scalar field. Each of Eqs. 7 and 8 expresses two different ways of writing the pushforward of a smooth vector field, acting on a scalar field, at points of M or $\mathbb{R}^4$, respectively. Thus, letting α and β be smooth vector fields on U, $p \in U$, and f be a smooth scalar field on V,

$$\phi_*([\alpha, \beta])(f)_{|p} = [\alpha, \beta](f \circ \phi)_{|\phi^{-1}(p)} \tag{9}$$

$$= \alpha(\beta(f \circ \phi))_{|\phi^{-1}(p)} - \beta(\alpha(f \circ \phi))_{|\phi^{-1}(p)} \tag{10}$$

$$= \alpha(\phi_*(\beta)(f) \circ \phi)_{|\phi^{-1}(p)} - \beta(\phi_*(\alpha)(f) \circ \phi)_{|\phi^{-1}(p)} \tag{11}$$

$$= \phi_*(\alpha)(\phi_*(\beta)(f))_{|p} - \phi_*(\beta)(\phi_*(\alpha)(f))_{|p} \tag{12}$$

$$= [\phi_*(\alpha), \phi_*(\beta)](f)_{|p}. \tag{13}$$

Equations 9 and 12 apply Eq. 7; Eqs. 10 and 13 apply Eq. 6; Eq. 11 applies Eq. 8. Since f and p were arbitrary, Eq. 13 holds generally. □

Before moving on, a note of comparison is in order regarding the sense in which approximate symmetries are observer-dependent. Applications of the SEP as discussed above are observer-dependent in the sense that they depend on the experimental apparatus available to measure "the strength of curvature effects" (Read et al., 2018, p. 17) that would distinguish solutions to an SR dynamical equation from those to a GR one in which curvature appears. These apparatus are also considered to be bounded, generally: "Whether you can detect tidal effects in a space the size of the room [you are in] depends on what kind of equipment you have access to, or in some cases how much time you have at your disposal!" (Brown, 2005, p. 170). By contrast, for approximate spacetime symmetries, there is no simple connection with curvature. On the one hand, it may be possible for curvature effects, insofar as they are coded in the second derivatives of the metric, to be quite strong in magnitude, yet the metric on the image of the local symmetry map is quite similar, so that their difference is quite small and does not preclude approximate symmetry for a fixed pair (h, ϵ). On the other, it may be possible for curvature effects to be quite small, yet for the differences in higher or lower (than the second) derivatives of the metric to be substantial enough to preclude approximate symmetry for (h, ϵ).

3.2 Poincaré Invariance of Equations and Poincaré Spacetime Symmetry

One question that arises naturally from the foregoing is what logical relationship Poincaré invariance of equations and Poincaré spacetime symmetries have with one another in relativistic spacetimes. In this subsection, I show that the one holding of a certain spacetime region does not imply the other on the same region. The examples are fairly simple and illustrate how the two concepts come apart.

3.2.1 Poincaré Spacetime Symmetry Without Poincaré Invariance of Equations

Here I adopt a simple example from Read et al. (2018, p. 24n51). It is special relativistic, thus set in Minkowski spacetime $(\mathbb{R}^4, \eta)$, which has not just approximate but exact Poincaré spacetime symmetry globally. However, it does not have Poincaré invariance of its dynamical equations in their simplest form.[9]

Consider a dust field with positive density ρ whose four-velocity field v^a forms a geodesic congruence. Its energy-momentum tensor is simply $T_{ab} = \rho v_a v_b$,

[9] Here I follow Read et al. (2018, p. 21) in construing one form to be simpler than another if it contains fewer terms. Although I am skeptical of the cogency of this notion—cf. my similar remarks about the hyperintensionality of minimal coupling in the concluding Sect. 5—insofar as it undergirds the definition and application of Poincaré invariance, which frames the question of its relation to Poincaré spacetime symmetry, I adopt it for those purposes without broader endorsement.

and its dynamical equations follow from the conservation condition $\nabla_a T^{ab} = 0$ (Malament, 2007, p. 243):

$$\nu^b \nabla_b \nu^a = 0, \tag{14}$$

$$\nu^b \nabla_b \rho + \rho(\nabla_b \nu^b) = 0. \tag{15}$$

The first, Eq. 14, is just the geodesic equation for the dust's worldlines. The second, Eq. 15, is a kind of energy conservation condition: its first term describes the change in energy density while its second the change in relative volume, both as described by a co-moving observer.

Thus it is not surprising that the equations of motion simplify in the relevant sense when expressed in coordinates well-adapted to such a co-moving frame. In particular, one can find such a frame in whose adapted coordinates all the spatial components of ν^a vanish, i.e., $\nu^\mu = 0$ for $\mu = 1, 2, 3$. Furthermore, the temporal component is constant, i.e., $\nu^0 = 1$ assuming normalization using geometric units. Thus Eqs. 14 and 15 simplify to a single equation:

$$\nu^0 \nabla_0 \rho = 0, \tag{16}$$

stating that the density is constant along any integral curve of ν^a. Clearly, the form of this equation is not preserved under Poincaré transformations of the chart in which it is expressed. Effectively, the four-velocity of the dust determines a preferred frame.

It might be objected that in order to count, the new field must not contribute to the energy-momentum tensor but rather be a part of spacetime structure or a fixed field of some sort.[10] In that case, one can introduce any timelike geodesic congruence τ^a separately to define a collection of preferred frames, and express the equations of motion of all matter fields in terms of it. Requiring that any dust field be spatially homogeneous with respect to coordinates adapted to these frames, which only differ by spatial rotations, results in the simplification of Eq. 15 to

$$\nu^0 \nabla_0 \rho + \rho(\nabla_\mu \nu^\mu) = 0. \tag{17}$$

For further examples, see Carroll and Lim (2004), Jacobson and Speranza (2015), and references therein.

[10] I take this to be a plausible reading of discussions in Brown (2005, p. 171) or Read et al. (2018, p. 24n52), which suggest some criterion like this as mandatory.

3.2.2 Poincaré Invariance of Equations Without Poincaré Spacetime Symmetry

The claim that it is possible to have Poincaré invariance of equations without even approximate Poincaré spacetime symmetry may seem at first in tension with the main results of Sect. 3.1, namely that every spacetime has approximate local Poincaré spacetime symmetry. One can resolve this tension through careful attention to the logical form of the definitions. Recall that a spacetime having a local symmetry just means that every point of the spacetime has a neighborhood that is the domain of a symmetry. But this does not entail that *every* neighborhood of a point is such a domain. Indeed, in a relativistic spacetime a generic neighborhood will not have even approximate Poincaré spacetime symmetry. Thus one way that a region could have approximate Poincaré invariance of equations applied within it but not approximate Poincaré spacetime symmetry would be if the approximation criteria for the former did not entail the latter. Whether this is so is not entirely clear because it is not clear what the relationship between fields in a spacetime approximately being solutions to a certain equation and that spacetime having approximate symmetries is. However, the discussion at the end of Sect. 3.1 about the mismatch between approximation of the metric and its derivatives by another and the small magnitude of curvature effects suggests a lack of entailment.

In any case, there are also topological constraints that can prevent a region from having Poincaré symmetry without precluding the Poincaré invariance of the equations for matter fields on that region. Consider again Minkowski spacetime, $(\mathbb{R}^4, \eta)$, and a Lorentz chart $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ whose domain is the whole manifold of the spacetime. For any dynamical equations for fields on it, if they are Poincaré invariant for the whole spacetime manifold, then they are so for any proper submanifold. So for any point of the manifold $p \in \mathbb{R}^4$, those equations are Poincaré invariant for the spacetime $(\mathbb{R}^4 - \{p\}, \eta_{|\mathbb{R}^4 - \{p\}})$. But that whole spacetime cannot have Poincaré spacetime symmetry, simply because as a spacetime region it has the wrong topology to support such symmetry. What were the Poincaré symmetry Killing vector fields on Minkowski spacetime are no long Killing vector fields, because any non-trivial flow along them is no longer well-defined. (They flow "into" the point removed.)

This can be the case even if the region considered is diffeomorphic to $\mathbb{R}^4$. Just as the dynamical equations that are Poincaré invariant on Minkowski spacetime are still Poincaré invariant on Minkowski spacetime sans an arbitrary point p, they are also Poincaré invariant on *any* region not containing p but with p in its closure. Such a region, considered now in the spacetime $(\mathbb{R}^4 - \{p\}, \eta_{|\mathbb{R}^4 - \{p\}})$, also cannot have Poincaré spacetime symmetry, for any putative Killing vector field on the region cannot be extended to a vector field that generates even a local one-parameter group of diffeomorphisms acting on the region.

Yet another further obstruction to Poincaré spacetime symmetry on a region arises when some putative Killing vector field on the region cannot be extended *smoothly* to a vector field that generates even a local one-parameter group of diffeomorphisms acting on the region. This can occur when the union of the images of the diffeomorphisms are not time- or space-orientable, even if their domain is.[11] For a simple two-dimensional example, consider the unit square with two of its boundary edges, $(0, 1) \times [0, 1]$, identifying the edges after a twist and removing the two resulting boundary points to form a Möbius band M (i.e., a Möbius strip without boundary). Label the now identified edges (sans two boundary points) as region E, and consider any flat Lorentz metric on M. The region $M - E$ is isometric to a simply-connected region of Minkowski spacetime. Thus, any dynamical equations for fields on that region are Poincaré invariant. But no putative set of Killing vector fields on $M - E$ generating Poincaré spacetime symmetries can be extended smoothly onto E.

4 Local Approximate Geometry

One of the examples from Sect. 3.2 sheds light on contrasting positions on the role of "local approximate spacetime geometry" in Torretti (1996) and Brown and Pooley (2001). To frame this contrast, recall that, in his interpretation of Schrödinger (1950), Brown (1997, p. 68) writes that "It is a fundamental assumption in GR that the local structure of spacetime, suitably defined, is special relativistic." Brown (1997, p. 71) has in mind in particular that "relative to local inertial frames (defined in the infinitesimal neighborhood of any event) all the laws of physics take on their special relativistic form. Put another way, the tangent space structure in GR is everywhere 'Lorentzian'." These paired claims, the first an expression (as discussed in Sect. 2) of the SEP and the second about spacetime geometry, form the basis for a putative explanation of SR's success (or, at least, how he takes this to be assumed in the formulation of GR).[12]

On one interpretation, this is what Torretti (1996, p. 136) argues, too, for according to him,

> no effect of gravity will be disclosed—within the agreed margin of precision—by any description of natural phenomena in terms of [the coordinates of a local Lorentz chart]; and that the laws of nature take the same form in [that neighborhood], when referred to [those coordinates], as they would referred to an ordinary Lorentz chart in a spacetime region where gravity is absent. . . . [This] implies that two experiments whose initial conditions

[11] Recall that a spacetime is time-orientable when there exists a continuous classification of timelike vector fields on the spacetime into future- and past-directed; it is space-orientable when there exists a continuous classification of orthogonal spacelike vector field triads on the spacetime into left- and right-handed (Wald, 1984, p. 60).

[12] Astute readers may wonder just how the second claim is supposed to be a paraphrase of the first. This will play a role in the emerging dispute and its resolution, below.

> read alike in terms of [local Lorentz charts] will also have the same outcome in terms of these charts.[13]

This corresponds with the first claim. For the second, he writes further that "the assumption that [the spacetime metric in GR] has the same signature as the Minkowski metric η is itself empirically motivated, insofar as it entails that every tangent space of the manifold is isometric with Minkowski spacetime, and thus accounts for the local success of Special Relativity" (Torretti, 1996, p. 139).[14] More precisely, "The Minkowski inner product on each tangent space induces—through the exponential mapping—a local approximate Minkowski geometry on a small neighborhood of each worldpoint. This accounts for the Lorentz invariance of the laws of nature referred to local Lorentz charts" (Torretti, 1996, p. 240). Thus, initially it seems that Brown and Torretti are in agreement regarding the explanatory nature of the paired claims and their status as assumptions, although Torretti does not equate the two claims but asserts rather that the first claim—about the laws of physics—follows from the second—about approximate local spacetime geometry.

However, Brown and Pooley (2001, p. 270) later deny this implication (hence implicitly Brown's earlier expression of the synonymy of these two expressions), insisting that laws involving the dynamics of matter fields must additionally satisfy the so-called minimal coupling condition for the former expression to be true:

> In our view, this claim [by Torretti] is a *non sequitur*. It is mysterious to us how the existence of a local approximate Minkowski geometry entails the Lorentz covariance of the laws of the non-gravitational interactions. Theories postulating a Lorentzian metric but which violate minimal coupling would involve non-Lorentz covariant laws. ... It seems to us that the local validity of special relativity in GR cannot be derived from what Torretti takes to be the central hypothesis of GR above, but must be independently assumed.

A matter theory is said to be minimally coupled when one arrives at the dynamical equations for its fields in GR by performing the following substitution procedure on its dynamical equations in SR: replace all instances of the Minkowski metric with a general Lorentzian metric, and all instances of the flat covariant connection with the Levi-Civita connection.[15] Thus, according to Brown and Pooley (2001), the first claim above—an expression of the SEP—is in fact independent of the second

[13] To make this last statement, Torretti (1996, p. 54) also invokes Einstein's "Principle of Relativity," that "The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or to the other of two Lorentz charts."

[14] It is important to distinguish Minkowski *spacetime*, which, having no preferred origin, has at most the structure of an affine space, from the tangent space of a relativistic spacetime at a point—what Hall (2004, p. 147) calls Minkowski *space*—which does have a preferred origin and thus has the structure of a vector space. Charitably, then, by "isometric with Minkowski spacetime" Torretti means either "isometric with that of any point of Minkowski spacetime" or, what is essentially equivalent, "isometric with Minkowski space."

[15] Actually, Brown (2005, p. 170) states that minimal coupling involves the non-appearance of terms depending on spacetime curvature in the dynamical equation, but Read et al. (2018, §3.2) rightly point out that this is in general not true, even accounting for the ambiguities of the application of the minimal coupling prescription as I have described it.

claim—about spacetime geometry—and so must be an additional assumption in the explanation of the success of SR.

Later, Brown (2005, p. 170–1) implicitly takes back the justification for this claim, that the reason the SEP does not follow from features of the local geometry of spacetime is that this geometry is compatible with non-minimally coupled matter fields, and non-minimally coupled matter fields have non-Lorentz invariant dynamical laws. He acknowledges instead that the dynamical equations for non-minimally coupled fields may still be Lorentz invariant. But then what sort of counterexample is supposed to confound Torretti's claim? Here is where the example of a non-Lorentz invariant matter theory on a relativistic spacetime comes in. Recall that this example was not invariant in the sense described precisely because there are a class of preferred frames in which the dynamical equations simplify. In other words, the definition of their simplest forms adverts to a spatiotemporal quantity or direction that is not derived from matter fields—it does not contribute to the energy-momentum tensor—and so must be a part of spacetime structure *beyond* the metric.

How does this example bear on Torretti's claim above, that the "local approximate Minkowski geometry on a small neighborhood of each worldpoint ... accounts for the Lorentz invariance of the laws of nature referred to local Lorentz charts" (Torretti, 1996, p. 240)? If by "laws of nature referred to Lorentz charts" Torretti meant merely laws that *use* Lorentz coordinates, this would indeed be a counterexample to his claim. However, in later writing Torretti (1999, p. 283n) clarifies his position with an example:

> the requirement of Lorentz invariance [imposed by SR] holds only for physical laws referred to a Lorentz chart. Therefore, the requirement cannot properly apply to Newton's laws, for the time variable that appears in them is not Einstein time. ... if one charitably replaces it with Einstein time, the Laws thus refurbished are not Lorentz invariant.

Here, "Einstein time" is simply the time variable that coordinatizes a Lorentz chart, while the time coordinate appearing in Newton's laws adverts to an observer-independent (i.e., observer-invariant) temporal structure. The other, spatial variables that appear in the laws, meanwhile, are those that normally coordinatize a Lorentz chart. Thus, in his (1996) Torretti was in fact referring to laws that advert *only* to Lorentz coordinates. Examples like the non-Poincaré invariant one from Sect. 3.2 are excluded because they refer to other spacetime structure that defines other coordinates. So, ultimately the criticism is evaded: that the local geometry of spacetime is approximately Minkowskian in the sense described does entail the Lorentz invariance of dynamical laws that only advert to the metric spacetime structure, hence only to Lorentz chart coordinates (Hall, 2004, pp. 286–7). It is after all these sorts of dynamical laws that are in play in SR, and the ones whose successful application is to be explained. The goal is not to recover (however approximately) laws not actually used in SR, but only those actually successfully applied; these laws are never excluded from Torretti's claim.

5 Conclusions, and Prospects for a Broader Explanation of the Success of SR

As I described in Sect. 1, the questions of the "local validity" of SR in GR ought to be interpreted as a question about reduction and explanation: GR reduces to SR locally, meaning that it explains the successful applications of SR within the bounds of allowable approximation. And one aspect of this explanation is the local application of spacetime symmetries, with their concomitant conserved quantities, in spacetimes that need have no global symmetries whatsoever. While the literature concerning the SEP may seem to address this question, I argued in Sect. 2 that in fact it does not; it focuses instead on the forms of dynamical laws and their invariance under Poincaré coordinate transformations, not Poincaré spacetime symmetries. I showed that neither of the two implies the other in Sect. 3.2, after describing in some detail my positive account of local approximate spacetime symmetries in Sect. 3.1, and how in particular every spacetime has approximate local Poincaré spacetime symmetry. Finally, in the previous section (Sect. 4), I applied these examples and a close reading of Torretti's account of the explanation of the success of SR to exonerate him from the charge of committing a non sequitur.

One issue raised in that last section was the role of "minimal coupling" in accounting for the invariance of the dynamical equations of matter fields under Poincaré transformations. Recall that minimal coupling is a prescription for generalizing an equation concerning a matter field in SR to one in GR. It is supposed to guarantee the Poincaré invariance of equations arising from it on some spacetime region (Read et al., 2018, Appendix A), yet, as shown at the end of Sect. 3.2, this does not even entail that there are approximate local spacetime symmetries on that region. So, it cannot be all that is needed to explain the success of SR matter theories, which requires showing that particular solutions to the dynamical equations—what experiments ultimately measure, after all—in some relevant sense approximate those of GR.[16] Moreover, that relativistic spacetimes have approximate local Poincaré symmetry does not guarantee that matter fields in GR behave locally in the relevant respects just like matter fields in SR. A full explanation of the local success of SR from GR—not just of the geometry of Minkowski spacetime but of matter theories formulated on it—will require something more.

There seem to be at least two components needed. First one must find a less "ambiguous" procedure for identifying the correct GR generalizations of SR dynamical equations for matter fields. Minimal coupling is considered "ambiguous" because it yields syntactically and semantically inequivalent results when applied to semantically equivalent but syntactically distinct equations. To say that two

[16] Cf. the demand of Sonego and Faraoni (1993, p. 1185) for a real scalar field satisfying the homogeneous screened Poisson equation: "We require that the *physical properties* of wave propagation—rather than the *form* of the wave equation—should reduce locally to those valid in flat spacetime. More precisely, we require that the physical features of the solutions be locally the same in both cases."

equations are semantically equivalent is just to say that they share the same intension. Yet two such (non-identical) equations can differ regarding whether they are minimally coupled equations. Thus, more properly, the problem with minimal coupling, particularly in its application to reduction, is its *hyperintensionality* (Nolan, 2014): it draws distinctions between cointensional properties (Hoffmann-Kolss, 2015). What justification does such a prescription have? I am skeptical that there is a justification or a "non-ambiguous" procedure in this sense—cf. the analogous problem for understanding how to quantize a classical theory (Feintzeig, 2017) or to generalize a spacetime theory to different dimensions (Fletcher et al., 2018). Perhaps therefore the problem is better approached the other way around: identify matter equations in GR that have matter equations of SR as a special case in Minkowski spacetime, then select among those the ones that have in fact been successfully applied.

Once one provides these candidates, one can then develop the second component: a better account of what it means for matter fields in a region of one spacetime to approximate those in a region of another. Here the methods of Fletcher (2014, 2018) applied in Sect. 3 to describe this for the spacetime metric and its derivatives ought to be of use. Although appeals to what one can detect with one's instruments here are surely correct, a more detailed explanation will surely be welcome.

Acknowledgments Thanks to Dennis Lehmkuhl, one of whose questions to me during the September, 2017 "Thinking About Space and Time" conference in Bern spurred the idea of this paper. Thanks also to James Read, Chris Wüthrich, and an anonymous referee for their comments, and especially to Jim Weatherall for suggesting a way to formalize a part of the proof of the main theorem and catching a serious error in a previous version thereof. Part of the research leading to this essay was written with the support from a Marie Curie Fellowship (PIIF-GA-2013-628533).

References

Brown HR (1997) On the role of special relativity in general relativity. International Studies in the Philosophy of Science 11(1):67–81

Brown HR (2005) Physical Relativity: Space-time structure from a dynamical perspective. Clarendon Press, Oxford

Brown HR, Pooley O (2001) The origin of the spacetime metric: Bell's 'Lorentzian pedagogy' and its significance in general relativity. In: Callender C, Huggett N (eds) Physics meets philosophy at the Planck scale: Contemporary theories in quantum gravity, Cambridge University Press, Cambridge, pp 256–272

Carroll S, Lim E (2004) Lorentz-violating vector fields slow the universe down. Physical Review D 70:123525

Ehlers J (1986) On limit relations between, and approximative explanations of, physical theories. In: Barcan Marcus R, Dorn GJW, Weingartner P (eds) Logic, Methodology and Philosophy of Science VII, Elsevier, Amsterdam, pp 387–403

Einstein A (1923) The foundation of the general theory of relativity. In: The Principle of Relativity, Methuen, pp 109–164, trans. Perrett W, Jeffery GB

Einstein A (1956) The Meaning of Relativity, 5th edn. Princeton University Press, Princeton, trans. Adams EP, Strauss EG, Bargmann S

Feintzeig BH (2017) On theory construction in physics: Continuity from classical to quantum. Erkenntnis 82(6):1195–1210
Fletcher SC (2014) Similarity and spacetime: Studies in intertheoretic reduction and physical significance. PhD thesis, University of California, Irvine
Fletcher SC (2016) Similarity, topology, and physical significance in relativity theory. The British Journal for the Philosophy of Science 67(2):365–389
Fletcher SC (2018) Approximate symmetry of spacetime structure, unpublished manuscript
Fletcher SC (2020) On representational capacities, with an application to general relativity. Foundations of Physics 50(4):228–249
Fletcher SC, Manchak J, Schneider MD, Weatherall JO (2018) Would two dimensions be world enough for spacetime? Studies in History and Philosophy of Modern Physics 63:100–113
Hall GS (2004) Symmetry and Curvature Structure in General Relativity. World Scientific, Singapore
Hoffmann-Kolss V (2015) On a sufficient condition for hyperintensionality. Philosophical Quarterly 65(260):336–354
Jacobson T, Speranza AJ (2015) Variations on an aetherial theme. Physical Review D 92:044030
Janssen M (2014) "No success like failure…": Einstein's quest for general relativity, 1907–1920. In: Janssen M, Lehner C (eds) The Cambridge Companion to Einstein, Cambridge University Press, Cambridge, pp 167–227
Malament DB (2007) Classical relativity theory. In: Butterfield J, Earman J (eds) Philosophy of Physics, Handbook of the Philosophy of Science, vol A, Elsevier, Amsterdam, pp 229–274
Nickles T (1973) Two concepts of intertheoretic reduction. The Journal of Philosophy 70(7):181–201
Nolan D (2014) Hyperintensional metaphysics. Philosophical Studies 171(1):149–160
Norton JD (1984) How Einstein found his field equations: 1912–1915. Historical Studies in the Physical Sciences 14(2):253–316
Norton JD (1993) General covariance and the foundations of general relativity: eight decades of dispute. Reports on Progress in Physics 56:791–858
Pauli W (1958) Theory of Relativity. Pergamon, trans. Field G
Read J, Brown HR, Lehmkuhl D (2018) Two miracles of general relativity. Studies in History and Philosophy of Modern Physics 64:14–25
Renn J, Sauer T (1998) Heuristics and mathematical representation in Einstein's search for a gravitational field equation. In: Goenner H, Renn J, Ritter J, Sauer T (eds) The Expanding Worlds of General Relativity, Einstein Studies, vol. 7, Birkhäuser, Boston, pp 87–125
Rosaler J (2019) Reduction as an *a posteriori* relation. The British Journal for the Philosophy of Science 70(1):269–299
Schrödinger E (1950) Space-time Structure. Cambridge University Press, Cambridge
Sonego S, Faraoni V (1993) Coupling to the curvature for a scalar field from the equivalence principle. Classical and Quantum Gravity 10(6):1185–1187
Torretti R (1996) Relativity and Geometry, Dover edn. Dover, Mineola, NY
Torretti R (1999) The Philosophy of Physics. Cambridge University Press, Cambridge
Wald RM (1984) General Relativity. University of Chicago Press, Chicago