Studies in History and Philosophy of Modern Physics 39 (2008) 102–153

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

Hilbert's 'Foundations of Physics': Gravitation and electromagnetism within the axiomatic method

K.A. Brading[a,*], T.A. Ryckman[b]

[a]*Department of Philosophy, University of Notre Dame, 100 Malloy Hall, IN 46556, USA*
[b]*Department of Philosophy, Stanford University, Stanford, CA 94305, USA*

Received 20 May 2007; received in revised form 14 August 2007

Thus all human cognition begins with intuitions, goes from there to concepts, and ends with ideas.
Kant, Critique of Pure Reason. (A702/B730)
Epigram to Hilbert (1899)

Abstract

In November and December 1915, Hilbert presented two communications to the Göttingen Academy of Sciences under the common title 'The Foundations of Physics'. Versions of each eventually appeared in the *Nachrichten* of the Academy. Hilbert's first communication has received significant reconsideration in recent years, following the discovery of printer's proofs of this paper, dated 6 December 1915. The focus has been primarily on the 'priority dispute' over the Einstein field equations. Our contention, in contrast, is that the discovery of the December proofs makes it possible to see the thematic linkage between the material that Hilbert cut from the published version of the first communication and the content of the second, as published in 1917. The latter has been largely either disregarded or misinterpreted, and our aim is to show that (a) Hilbert's two communications should be regarded as part of a wider research program within the overarching framework of 'the axiomatic method' (as Hilbert expressly stated was the case), and (b) the second communication is a fine and coherent piece of work within this framework, whose principal aim is to address an apparent tension between general invariance and causality (in the precise sense of Cauchy determination), pinpointed in Theorem I of the first communication. This is not the same problem as that found in Einstein's 'hole argument'—something that, we argue, never confused Hilbert.

*Corresponding author.
E-mail addresses: kbrading@nd.edu (K.A. Brading), tryckman@stanford.edu (T.A. Ryckman).

1355-2198/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsb.2007.08.002

When citing this paper, please use the full journal title *Studies in History and Philosophy of Modern Physics*

1. Introduction

In November and December 1915, Hilbert gave two presentations to the Royal Göttingen Academy of Sciences under the common title 'The Foundations of Physics'. Distinguished as 'First Communication' (Hilbert, 1915b) and 'Second Communication' (Hilbert, 1917), two papers (or 'notes', as they are widely known) eventually appeared in the *Nachrichten* of the Academy.[1] The First Communication, which quickly entered the canon of classical general relativity, has recently become the object of renewed scholarly scrutiny since the discovery of a set of printer's proofs (Hilbert, 1915a) dated 6 December 1915 (henceforth the 'Proofs').[2] With the exception of Renn and Stachel (1999), the Second Communication has not been given the same detailed reconsideration. However, the analysis of the Second Communication by Renn and Stachel seriously misrepresents its aims, content, and significance, and also its links to the First Communication. Our aim in this paper is to show that Hilbert's Second Communication is a natural continuation of his First Communication, that it contains important and interesting further developments of that project, and above all that it sheds needed illumination on Hilbert's assessment of the epistemological novelty posed by a generally covariant physics.

Hilbert's notes on 'Foundations of Physics' traditionally have been assessed solely in terms of the contributions they made to general relativity, as that theory is known in its completed form.[3] From this vantage point, they present a mixed record of achievement, ranging from genuine insight (the Riemann scalar as the suitable invariant for the gravitational action) through incomprehension (Hilbert's interpretation of electromagnetism as a consequence of gravitation) to abject failure (attachment to the untenable electromagnetic theory of matter of Gustav Mie). The usual implication is that Hilbert's principal intent in November 1915 was to arrive at a theory of gravitation based on the principle of general covariance in one blinding flash, masterfully wielding an arsenal of axiomatized advanced mathematics. Thus arose 'the legend of a royal road to general relativity' (Renn & Stachel, 1999, p. 1) through the axiomatic method, whilst Hilbert's reputed remark that 'physics is much too difficult for physicists' has been widely understood to epitomize a haughty mathematical arrogance (Reid, 1970, p. 127). Correspondingly, some historians of general relativity have concluded that Hilbert 'attached a kind of metaphysical significance to variational methods' (Rowe, 1999, p. 201), while others have regarded Hilbert's approach as evincing an optimistic demonstration of the Göttingen-based ideology of a 'pre-established harmony' between mathematics and

[1] English translations of these papers, as well as of Hilbert (1915a), are now available in Renn & Schemmel, Eds. (2007). Unless otherwise noted, all translations in this paper are our own. For readability, equations have been renumbered as necessary.

[2] Corry, Renn, & Stachel (1997); see e.g., Rowe (1999, 2001), Renn & Stachel (1999), Stachel (1999), Sauer (1999), Vizgin (2001), Corry (2004), and Sauer (2005).

[3] A welcome exception is Corry (2004), who treats Hilbert's notes as part of his program for the axiomatization of physics.

physics (Pyenson, 1985).[4] Finally, some have relished pointing out where Hilbert's elaborate mathematical constructions were either inadequate to the complexities of the initial value problem in general relativity[5] or simply led to hopelessly failed physics (Stachel, 1992; Renn & Stachel, 1999, pp. 77, 81–83).

Our contention is that viewing Hilbert's notes solely in terms of contributions made to general relativity as that theory is canonically understood, radically occludes internal motivations, which are largely logical and epistemological, and so casts them in a misleading light. In so doing, the explicitly stated epistemological intent of the 'axiomatic method' is willfully ignored, as are Hilbert's own express assertions regarding his construction as a triumph of that method. Although understandable in terms of the intellectual small change of 'textbook' histories, such accounts overlook or downplay fundamental philosophical and methodological differences with Einstein, in emphasis as well as in detail, concerning the significance of general covariance, arguably impeding clarification on that vexed issue for decades. But set within the logical and epistemological context of the 'axiomatic method', Hilbert's two notes may be seen to have the common goal of pinpointing, and then charting a path toward resolution of, the tension between causality and general covariance that, in the infamous 'hole argument', had stymied Einstein from 1913 to the autumn of 1915.[6] Unlike Einstein's largely informal and heuristic extraction from the clutches of the 'hole argument', Hilbert stated the difficulty in a mathematically precise manner as an ill-posed Cauchy problem in the theory of partial differential equations, and then indicated how it can be resolved. As we will show, material cut from the proofs establishes this essential thematic linkage between the two notes and redeems Hilbert's claim that tension between causality and general covariance, precisely formulated in Theorem I of the First Communication, was the 'point of departure' for his axiomatic investigation.

Einstein and Hilbert were engaged in qualitatively different enterprises that only partially overlapped. In contrast to Einstein, Hilbert's goals were at least as much logical and epistemological, according to the character of the axiomatic method, as they were physical. We concur with the judgment of Felix Klein, who wrote, in 1921, that 'there can be no talk of a question of priority, since both authors pursued entirely different trains of thought (and to be sure, to such an extent that the compatibility of the results did not at

[4]To be sure, Hilbert occasionally voiced such sentiments in his lectures; e.g., noting the simplicity of the Maxwell equations in four-dimensional formulation, and how appeal to the simplest differential invariants in Einstein's theory of gravitation yielded the accurate correction of Newton's theory (regarding the precession of Mercury's perihelion), Hilbert remarked that such results gave 'an impression of pre-established harmony. We confront here the remarkable fact that apparently matter entirely obeys the formalism of mathematics. There appears here a previously unsuspected agreement between being (*Sein*) and thought that we must provisionally accept as a miracle' (Hilbert, 1919–1920, p. 69). Hilbert's acceptance was indeed provisional; a central concern in the remainder of these lectures is to analyze and explain the 'miracle' from what would later be termed 'the finite point of view' (*die finite Einstellung*). On the latter, see Section 8 below.

[5]As will be seen, Hilbert's main concern is with the Cauchy problem of evolving the initial data forwards. Hilbert also shows some concern for the problem of finding a suitable initial value hypersurface, but the problems associated with then ensuring that the initial data, specified on such a surface, are consistent with the field equations, have yet to become apparent (and will take some time to emerge in the study of Einstein's general theory of relativity). These problems are, of course, all related to one another (see Appendix A).

[6]See Norton (1984, pp. 286–291; 1993, § 1–3), Stachel (1993), and Ryckman (2005, § 2.2.2), for presentation and discussion of the 'hole argument', and for additional references.

once seem assured)'.[7] For Hilbert, the principal outcomes arrived at by the axiomatic method concern his revisiting the principle of causality, and his revisions of Kantian epistemology, in the light of generally covariant physics.

The structure of our paper is as follows.

In Section 2 we present what we call 'the essential context': Hilbert's axiomatic method and its presupposition of central tenets of Kantian epistemology.

In Section 3, we briefly review the published version of Hilbert's First Communication, proceeding in Section 4 to emphasize differences in content between the December Proofs and the published version. Since this subject has been extensively treated in Sauer (1999), in Renn and Stachel (1999), and more recently in Corry (2004), our treatment will highlight only the central features, omitting many details that can be found in these sources. We shall see that a passage cut from the proofs elucidates the problem pinpointed by Theorem I, and this, we claim, provides essential thematic linkage to Hilbert's Second Communication (see Section 6). Section 5 returns to the topic of the axiomatic method, and examines the aims and achievements of this method as it appears in the First Communication.

Section 6 concerns Hilbert's Second Communication. Following a brief introduction, we review the secondary literature, and then turn our attention to an exposition of the content of the Second Communication. We see how Hilbert now sought to resolve the challenge posed by Theorem I—the tension between general covariance and causality. We show that Hilbert's employment of the axiomatic method identified an epistemological novelty emerging in generally covariant physics regarding the *constitution of physical objectivity* as this is understood in a broadly transcendental idealist sense, most prominently displayed in his subsequent remarks regarding the axiom of general invariance. In further addressing the related matter of the vexing problem of causality in the new physics of general covariance, deemed a 'pseudo-geometry' and not, as before, as field physics set within a 'background geometry', Hilbert sought to remove all global 'pseudo-Euclidean' presuppositions, equivalent to 'action-at-a-distance'.

Armed with our new understanding of Hilbert's 'problem of causality', Section 7 explains why this is not the same problem as Einstein faced in his 'hole argument'.

In Section 8, we look beyond Hilbert's 1915 and 1917 papers to consider his further reflections on the epistemological significance of the respective principles of general covariance and causality and the ground of Hilbert's subordination of the latter to the former. We explain in detail the revisions of Kant that Hilbert believed were required in the face of the new generally covariant physics.

2. The essential context: Hilbert's axiomatic method and Kantian epistemology

There are two pieces of context that we believe are crucial to correctly understanding Hilbert's treatment of generally covariant physics: his axiomatic method, and his appeal to Kantian epistemology. These themes infuse our discussion throughout.

Hilbert's First Communication opens with a declaration that his investigation of the foundations of physics is undertaken 'in the sense of the axiomatic method' ('*im Sinne der*

[7] '*Von einer Prioritätsfrage kann dabei keine Rede sein, weil beide Autoren ganz verschiedene Gedankengänge verfolgen (und zwar so, daß die Verträglichkeit der Resultate zunächst nicht einmal sicher schien*)'. This remark occurs in a note (p. 566, n. 8) added to the 1921 reprint of Klein (1917).

axiomatischen Methode'), and it concludes with the striking claim that the results he has obtained redound 'certainly to the most magnificent glory of the axiomatic method'. Unless these passages are mere rhetorical embellishment, they establish that the 'axiomatic method' (whatever that may be) played an integral part in Hilbert's work on the foundations of physics. It is our contention that understanding the significance of Hilbert's setting his results squarely within the frame of 'the axiomatic method' is essential for correctly interpreting his First and Second Communications.

What, then, is 'the axiomatic method'? Einstein himself appears to have been somewhat skeptical regarding Hilbert's claims of the method's intended significance, placing the term in scare quotes in a notably sarcastic aside to Weyl.[8] In the literature, it has been widely, if tacitly, assumed that Hilbert's references to 'axiomatic method' simply signal the derivation of his 14 fundamental field equations, as well as several subsidiary theorems, from two principal axioms.[9] However, in Hilbert's usage this term implicates not merely a typical mathematical concern with the rigorous explicit statement of a theory, but rather also connotes a specifically *logical and epistemological* method of investigation for 'deepening the foundations' of the theory. Hence, by invoking 'the axiomatic method', Hilbert was calling attention to a specifically epistemological method of investigation of mathematical theories (including those of physics) that he pioneered, and which he saw as closely tied to the nature of thought itself.[10]

Any attempt to understand attribution of epistemological significance to the axiomatic method must begin with Hilbert's attitude toward geometry, which Hilbert always regarded as a *physical* science (indeed, the paramount physical science), and which served as a model for his treatment of physical axioms (Hallett & Majer, 2004, p. 66). In published articulation, the 'axiomatic method' debuted in Hilbert's classic Gauss–Weber *Festschrift* essay, *Grundlagen der Geometrie* (1899). The epigraph to Hilbert's essay has been little noticed, yet is worth quoting in the original German, for it is Kant's most concise statement (see the discussion in Section 8) of how cognition arises from the distinct sources of intuition, concepts, and ideas:

> *So fängt denn alle menschliche Erkenntnis mit Anschauung an, geht von da zu Begriffen und endigt mit Ideen* (A702/B730).

To consider the appropriateness of this passage, recall that in *Grundlagen der Geometrie*, Hilbert presented a rigorous axiomatization of Euclidean geometry, beginning from the famous initial posit ('*Wir denken uns*…') of a domain of three non-descript systems of 'things' (*Dingen*) which he termed 'points', 'straight lines', and 'planes'. Of course, each term (and the relations each enters into with the others) has a sense familiar from our everyday experience of objects, and so empirical intuition supplies the basic facts of

[8]Einstein to H. Weyl, 23 November 1916: 'Certainly I'll admit that finding the *suitable* (*geeigneten*) hypothesis, respectively, Hamiltonian function [i.e., Lagrangian density], for the construction of the electron forms one of the most important contemporary tasks of theory. But the "axiomatic method" can be of little help with this (*kann dabei wenig nützen*)' (Einstein, 1998, p. 366).

[9]E.g., Guth (1970, p. 84), Mehra (1974, pp. 26, 72, n. 145), Wightman (1976, p. 153), Pais (1982, p. 257): 'Suffice it to say that it was Hilbert's aim to give not just a theory of gravitation but an axiomatic theory of the world'. As we will see, the December Proofs contain three axioms.

[10]Hallett (1994, p. 162) quotes from Hilbert's 1905 Summer Semester Lectures '*Logische Principien des mathematischen Denkens*', 'The general idea of [the axiomatic method] always lies behind any theoretical and practical thinking'.

geometry subjected to the axiomatic treatment. In point of fact, Hilbert regarded the axiomatization as 'the logical analysis of our spatial intuition'.[11] But for the purposes of such an analysis, meanings of these terms are neither antecedently assumed nor primitively defined; rather the terms are 'implicitly defined', i.e., such meaning as accrues to each term within the axiomatic structure is acquired through the logical relations it enters into by virtue of its occurrence in any of the five classes of axioms and in all ensuing theorems. Accordingly, these geometric axioms compactly 'express certain interrelated fundamental facts of our intuition'.

In more general terms, and as Kant's directive prescribes, the axiomatic method is conceived as a logical analysis that begins with certain 'facts' presented to our finite intuition or experience. Both pure mathematics and natural science alike begin with 'facts', i.e., singular judgments about 'something … already … given to us in representation (*in der Vorstellung*): certain extra-logical discrete objects that are intuitively present as an immediate experience prior to all thinking'.[12] As the axiomatic method is characterized in Hilbert's Göttingen Winter Semester lectures in 1922/1923, analysis then determines the concepts under which such given facts can be classified and arranged, and next attempts to formulate the most general logical relations among these concepts, a 'framework of concepts' (*Fachwerk von Begriffen*) crowned with the fewest possible number of principles. These axioms are, as far as possible, independent of the particular intuitions (and so, concrete facts) from which the process started (see immediately below). But in addition, by subjecting the intuitively given data to logical analysis, the axiomatic method is concerned to separate out and highlight the self-sufficiency of the mathematical subject matter (which may then be developed autonomously), quite apart from any particular reference associated with particular terms or relations. In this way, a separation is effected between logical/mathematical vs. intuitional/experiential thought, even as the latter has thus been arranged in deductive form. Indeed, it is just 'the service of axiomatics'

> to have stressed a separation into the things of thought (*die gedanklichen Dinge*) of the (axiomatic) framework and the real things of the actual world, and then to have carried this through.[13]

When applied to any theory covering a sufficiently known domain of facts, whether of mathematics or natural science, the axiomatic method is a procedure of finding, for any given proposition of the theory, the premises from which it follows. The epistemological orientation of such a method is obvious, and indeed, it rigorously implements the more general epistemological approach of regressive or analytic methods for isolating and determining the most general basic propositions on which rest a given body of knowledge.[14] In each case, the aim is not, at least in the first instance, the discovery or

[11]'*Die bezeichnete Aufgabe läuft auf die logische Analyse unserer räumlichen Anschauung hinaus*' (Hilbert, 1899, p. 3).

[12]Hilbert (1922, p. 161, English trans., p. 1121). Of course, for Hilbert, the basic objects of number theory, the positive integers, or rather the *signs* that are their symbolic counterparts, are given in a quasi-spatial, but not in *spatial* or *temporal*, intuition.

[13]Hilbert Winter Semester 1922/1923 lectures *Wissen und mathematisches Denken.* Ausgearbeitet von Wilhelm Ackermann. Mathematische Institut Göttingen. Published in a limited edition, Göttingen, 1988; as translated in Hallett (1994, p. 167).

[14]This theme is taken up by Leonard Nelson (1928), an exploration of the 'points of contact between critical (i.e., Kantian) philosophy and mathematical axiomatics' (in Hilbert's sense). In a letter of 30 July 1918 (cited and

recognition of *new* laws or principles, but the conceptual and logical clarification or reconstruction of known ones (cf. Majer, 2001, p. 19). Finally, and as its culmination, the axiomatic method is concerned to demonstrate that the axioms of the theory thus selected possess three meta-logical properties or relations: of mutual consistency, independence, and completeness.[15] Combining all these aspects together, successful pursuit of the axiomatic method leads to a 'deepening of the foundations' (*Teiferlegung der Fundamente*), i.e., of the *mathematical foundations*, of any theory to which it is applied, and this, indeed, is the overall objective.[16]

Two further considerations require emphasis. First, a theory axiomatized according to the axiomatic method satisfies, according to Hilbert, the criteria of existence and truth solely through a *consistency proof*, i.e., a demonstration of the mutual consistency of the axioms and all their consequences. This was Hilbert's view already in *Grundlagen der Geometrie* (again, we recall that Hilbert always regarded geometry as a natural science) when it became a well-known bone of contention with Frege (e.g., Corry, 2004, pp. 112–114). Yet the axiomatic method requires still more: that consistency obtain not only with respect to the various axioms, but also (see below) with respect to the 'conditions of possibility of all conceptual knowledge and all experience'. In other words, all appearance of conflict between the different contributions to scientific knowledge—intuitions, concepts, ideas—should be removed, yielding a 'complete agreement and most pleasant harmony' between the experiences of everyday life and 'the most demanding sciences'.[17] This emphasis on the compatibility between the different sources of knowledge is crucial for understanding Hilbert's project in the Second Communication (see Section 6, below).

Secondly, the mathematical axioms standing at the pinnacle of the *Fachwerk von Begriffen* are not only general but also *ideal*: more precisely, they are regarded as 'ideas' in Kant's regulative sense, i.e., principles or 'rules of possible experience' possessing an 'objective but indeterminate validity' (A663/B691) but not a constitutive employment in cognition (however, we will see in Section 8 that Hilbert's revision of the Kantian account of physical objectivity rejects a sharp *constitutive/regulative* distinction). According to the

(*footnote continued*)
translated in Peckhaus, 1994, p. 104), Hilbert wrote to the Prussian Education Minister of his wish 'above all to propagate the connections between mathematics and philosophy', naming as allies in this regard 'among philosophers … Husserl and Nelson [as] the two most prominent personalities, and to my mind, it is no accident that these two had appeared on the mathematical soil of Göttingen'.

[15]Hilbert's 1905 Summer Semester Göttingen lectures '*Logische Prinzipien des mathematischen Denkens*' already characterized the general idea of the axiomatic method as stressing the consistency, independence, and completeness of an axiom system. See Peckhaus (1990), p. 59.

[16]Hilbert (1918, p. 407; English translation, p. 1109): 'The procedure of the axiomatic method, as it is expressed here, amounts to a *deepening of the foundations* of the individual domains of knowledge, just as becomes necessary for every edifice that one wishes to extend and build higher while preserving its stability'.

[17]Lecturing in Summer Semester 1921 on the 'Basic Ideas of Relativity Theory' (*Die Grundgedanken der Relativitätstheorie*), Hilbert stressed that the new conceptions of space, time, and motion of Einstein's theory were still compatible with 'the traditional intuition' of 'everyday life, our practice and custom': 'Thus we have listed all the essential features of the old conception of space, time, and motion. But … it is still absolutely necessary to bring to mind how excellent this conception of spacetime has proved to be. As far as natural sciences and their applications are concerned, we find that everything is based on this conception. And in this construction everything fits together perfectly. Even the boldest speculations of physicists and astronomers are brilliantly confirmed in the minutest detail so that one can say that the experiences of everyday life, our practice and custom, the traditional intuition and the most demanding sciences were in complete agreement and most pleasant harmony with each other'. As cited and translated in Majer (1995, p. 274).

axiomatic method, in virtue of their ideality, and so severance from experience and intuition, axioms can play at best a hypothetical role in cognition.

Perhaps Hilbert's last published statement of his epistemological credo occurred in a 1930 paper entitled 'Knowledge of Nature and Logic'. There, in the course of a discussion of how modern science has led to the judgment that Kant had far overestimated the role and extent of *a priori* elements in cognition, Hilbert nonetheless endorsed a conception of such elements as 'nothing more and nothing less than a basic point of view (*Grundeinstellung*) or expression for certain unavoidable preconditions of thinking and experience'.[18] He concluded that what remains of Kant's synthetic *a priori* is just this 'intuitive *a priori* point of view' that is presupposed in all theoretical concept construction in mathematics and physics. But Hilbert stressed that this was in full agreement with the basic tendency of Kantian epistemology:

> Thus the most general and fundamental idea of Kantian epistemology retains its significance: namely, the philosophical problem of determining that intuitive *a priori* viewpoint (*jene anschauliche Einstellung a priori*), and thereby of investigating the conditions of the possibility of all conceptual knowledge and of all experience.[19]

We discuss Hilbert's own modifications of Kantian epistemology in Section 8, below. Now we turn to the details of Hilbert's First and Second Communications, and we return to the topic of the axiomatic method in the context of physics in Section 5.

3. Hilbert's First Communication on 'The Foundations of Physics' (published version)

According to the annotation on the published version of the paper, Hilbert's First Communication was presented at the 20 November 1915 session of the Royal Göttingen Academy of Sciences. Traditionally, the date of submission was the only date appearing on publications in the *Nachrichten* of the Academy (Rowe, 2001, p. 418). However, with the discovery of the December Proofs in 1993 it was learned that the version submitted on 20 November differs considerably from that appearing in the published *Nachrichten* on 31 March 1916.[20] In this section we outline the content of the First Communication, and examine what was cut from the December Proofs. While this issue has been considered in detail before (Sauer, 1999; Renn & Stachel, 1999; Vizgin, 2001), our purpose is rather different. Our interest lies in comparing the content of the First Communication, and especially what was cut from the Proofs, with what was published as the content of the Second Communication. This enables us to see that the Second Communication treats in

[18] Hilbert (1930, p. 961).

[19] While the intent of these remarks on Kant is apparently to emphasize Hilbert's conviction that 'outside of deduction and experience, there is still a third source of cognition (*Erkenntnisquelle*)', it is not particularly clear from the text what is meant by 'intuitive *a priori* viewpoint'. But elsewhere Hilbert describes this intuitive viewpoint (*anschauliche Einstellung*) as 'an *a priori* insight … that the applicability of the mathematical way of reflection over the objects of perception is an essential condition for the possibility of an exact knowledge of nature', an epistemological position, Hilbert goes on to state, that 'seems to me to be certain' ('*Wissen und mathematische Denken*', Göttingen Winter Semester 1922/1923 lectures. Ausgearbeitet von W. Ackermann. Mathematische Institut Göttingen. Published in a limited edition, Göttingen, 1988. As cited and translated in Corry, 2004, p. 429).

[20] Hilbert was sending offprints to colleagues in mid-February 1916; see Sauer (1999, p. 543, n. 74).

detail an issue raised in the Proofs of the First Communication, but whose resolution Hilbert was forced to revisit.

3.1. Hilbert's aim

As legend has it, in November 1915, Hilbert engaged in a competition with Einstein to arrive at the generally covariant field equations of gravitation. Certainly, there was some sort of a 'race': no other term quite so well suits the frenzied activities of Einstein and Hilbert in that month. But this can by no means have been Hilbert's only aim. In seeking a derivation of the field equations of gravitation from a variational principle, Hilbert upped the ante in postulating a single generally invariant 'world function', a Lagrangian for both the gravitational *and* the matter fields, from which the fundamental equations of a pure field physics might be derived. In astonishing testimony to his belief in the axiomatic method's power to 'deepen the foundations' of a theory, this objective is stated as the main aim in both published versions of Hilbert's two communications, and indeed is still posed as late as 1923 (Hilbert, 1915b, p. 395; 1917, pp. 63–64; 1923, pp. 12–13).

The First Communication accordingly begins with a declaration that the investigations of Einstein and Mie have 'opened new paths for the investigation of the foundation of physics'. Hilbert announced that his aim is to set up '*in the sense of the axiomatic method*' (*im Sinne der axiomatischen Methode*—our emphasis) a new system of fundamental equations of physics on the basis of two (or, three, in the Proofs) axioms 'of ideal beauty', encompassing in a single theory both Einstein's theory of gravitation and Gustav Mie's theory of matter.[21] These two theories were, in 1915, clear candidates to be the fundamental theories of physics.[22] Expressing Einstein's theory of gravitation in terms of the 10 independent gravitational 'potentials' $g_{\mu\nu}$, and providing a generally invariant generalization of Mie's theory expressed in terms of the 4 electrodynamic potentials q_s, Hilbert employed highly sophisticated mathematical techniques to draw out the consequences of his two principal axioms, as we will see in more detail below. While Hilbert's ambitious maximum goal was neither attained nor attainable (solutions to the non-linear generalized Maxwell equations were found to be physically untenable, i.e., not corresponding to the particulate structure of matter), it is clear that Hilbert was nonetheless extremely pleased with the outcome of his application of the axiomatic method to conjoin the two theories. The triumphal language at the end of his First Communication can be understood as expressing Hilbert's great satisfaction with the illumination gained in revealing unsuspected mathematical relations between the field equations for gravitation

[21]See in particular the discussion of Mie's theory in Corry (2004, pp. 299–315). Both Corry and Sauer (1999) emphasize that Hilbert's knowledge of Mie's theory was in the form given to it by Born (1914).

[22]Einstein's antipathy toward Mie's theory of matter is well known. It also anteceded Hilbert's reformulation of Mie's theory. Already in mid-August 1913 in a letter to Erwin Finlay Freudlich, Einstein referred to the Mie theory as 'fantastic', remarking, 'in my opinion, it has only a vanishing small inner probability' (Einstein, 1993, p. 550). However, Sauer (2002, p. 231) notes (in response to related claims by Renn and Stachel as to the Mie theory's implausibility in 1915): 'I do not think that the electromagnetic world view was unambiguously outdated at the time and that you could not have had well-founded reasons to believe that speculations along the lines of Mie's theory would give you a reasonable theory of matter'. Indeed, the second (1923) edition of Von Laue's widely used treatise on general relativity still contains a 5-page section (§29) on Mie's theory of matter *in the context of Einstein's theory*.

and for electrodynamics. In what follows we sketch how this illumination was achieved in his First Communication.

3.2. Schematic outline

The core of Hilbert's approach lies in two axioms, which he states immediately after some preliminary definitions.

Axiom I. ('Mie's Axiom of the World Function'). Hilbert proposed a variational argument formulated for a 'world function' H,[23] depending on the 10 gravitational potentials $g_{\mu\nu}$, their first and second derivatives, as well as the 4 electromagnetic potentials q_s, and their first derivatives:

$$\delta \int H\sqrt{g}\,\mathrm{d}\omega = 0 \quad (g = \det|g_{\mu\nu}|,\ \ \mathrm{d}\omega = \mathrm{d}w^1\,\mathrm{d}w^2\,\mathrm{d}w^3\,\mathrm{d}w^4). \tag{1}$$

Axiom II. ('Axiom of General Invariance'). The world function H is an invariant with respect to arbitrary transformations of the 'world parameters' w_s ($s = 1,2,3,4$).

Hilbert's use of the term 'world parameters' in place of the standard locution 'spacetime coordinates' is instructive. As expressly stated in his Second Communication, and as Mie noted that same year,[24] it is intended to highlight the analogy Hilbert sought to draw between the arbitrariness of parameter representation of curves in the calculus of variations, and the arbitrariness of coordinates on a spacetime manifold. Hilbert was, of course, a grand master of the calculus of variations, as this communication will demonstrate. In both cases, objective significance will accrue only to objects invariant under arbitrary transformation of the parameters, respectively, coordinates. As Hilbert used precisely the same language of 'world parameters' also in the Proofs, this is *prima facie* evidence that his views regarding the lack of physical meaningfulness accruing to spacetime coordinates were already in place. Similarly, in both versions of the First Communication Hilbert affirms that this axiom is

> the simplest mathematical expression for the demand that the interconnection of the potentials $g_{\mu\nu}$ and q_s is, in and for itself, completely independent of the way in which one designates the world points through world parameters (Hilbert, 1915a, p. 2; 1915b, p. 396).

Anticipating our later discussion (in Section 7) of Hilbert and the 'hole argument', we note that in the 1924 republication of Hilbert's two notes in *Mathematische Annalen,* the term

[23] The term 'world function', while appearing in Mie (1912–1913) and reminiscent of Minkowski, was used by Hilbert already in his 1905 lectures on Newtonian continuum mechanics; see Corry (2004, p. 152).

[24] Hilbert (1917, p. 61): 'Just as in the theory of curves and surfaces an assertion for which the parameter representation of the curve or surface has been chosen has no geometric meaning for the curve or surface itself, so we must also in physics designate an assertion as *physically meaningless* (*physikalisch sinnlos*) that does not remain invariant with respect to arbitrary transformation of the coordinate system'. Mie (1917, p. 599) also stressed this analogy.

'world parameters' has been dropped while the sentence has been reformulated explicitly in terms of the physical meaninglessness of spacetime coordinates:

> Axiom II is the simplest mathematical expression for the demand that the coordinates in themselves have no manner of physical meaning, but rather represent only an enumeration of the world points in such a way as is completely independent of the interconnection of the potentials $g_{\mu\nu}$ and q_s (Hilbert, 1924, p. 4).

Nonetheless, given what is surely a semantic equivalence between the two sentences, we cannot agree with Corry's assessment that this change ('Hilbert now added a paragraph') represents a change 'distancing [Hilbert] from the position that was variously insinuated in his earlier versions' (Corry, 2004, p. 401).

Before proceeding further Hilbert then stated, without proof, a theorem described as the '*Leitmotiv* of my theory', whose content may be more briefly stated thus:

Theorem I. ('*Leitmotiv*'). In the system of n Euler–Lagrange differential equations in n variables obtained from a generally covariant variational integral such as in Axiom I, 4 of the n equations are always a consequence of the other $n-4$ in the sense that 4 linearly independent combinations of the n equations and their total derivatives are always identically satisfied.[25]

One of Hilbert's principal claims is that, as a consequence of Theorem I, electromagnetic phenomena may be regarded as consequences of gravitation. We discuss this claim in Section 3.3 below. The theorem also gives rise to Hilbert's 'problem of causality', which we discuss in Section 4.2.

Hilbert next turns to the derivation of the Euler–Lagrange differential equations from his invariant integral, by differentiation of H with respect to the $g_{\mu\nu}$ and their first and second derivatives. This yields (equations (4) and (5) in Hilbert's numbering), on the one hand, ten equations for the gravitational potentials,

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k\,\partial w_l}\frac{\partial\sqrt{g}H}{g_{kl}^{\mu\nu}} = 0, \tag{2}$$

or, in Hilbert's abbreviation,

$$[\sqrt{g}H]_{\mu\nu} = 0\left[g_l^{\mu\nu} = \frac{\partial g^{\mu\nu}}{\partial w_l};\ \ g_{lk}^{\mu\nu} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l\,\partial w_k}\right],$$

while, on the other, differentiation of H with respect to the electromagnetic potentials q_s and their first derivatives yields four equations.[26]

$$\frac{\partial\sqrt{g}H}{\partial q_h} - \sum_\sigma \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}} = 0, \tag{3}$$

[25] Hilbert (1915a, pp. 2–3; 1915b, p. 397). Later in the paper, Hilbert regards the invariant H as the additive sum of *two* general invariants $H = K+L$ (see Section 4.1 below), where K represents the source-free gravitational Lagrangian and L is the source term associated with the addition of matter fields (the electromagnetic field in Hilbert's theory). As Klein (1917, p. 481) first pointed out, there are therefore eight identities available; four associated with K and four with L. According to Klein, the identities associated with L reveal that the conservation laws of the matter field equations are consequences of the gravitational field equations, and he concluded that they therefore 'have no physical significance'. This redundancy in the field equations, a feature of the generally invariant structure of the theory, prompted Hilbert's interpretation of the electromagnetic equations as a consequence of the gravitational equations, as discussed in Section 3.3.

[26] The form of equations (2) and (3) is trivially algebraically different between the Proofs and the published version. Here we follow the published version. For ease of comparison with the text, we also follow Hilbert's

or

$$[\sqrt{g}H]_h = 0 \quad \left[q_{hk} = \frac{\partial q_h}{\partial w_k} \ (h, k = 1, 2, 3, 4)\right].$$

The 14 equations (2) and (3) are termed, respectively, 'the basic equations of gravitation, and electrodynamics or generalized Maxwell equations'. On the assumption that the Mie theory rendered a viable theory of matter, these equations encompass the entirety of fundamental physics.

The remainder of the paper concerns Hilbert's treatment of energy, which includes his demonstration of a connection between the phenomena of gravitation and of electromagnetism. We turn to this issue now.

3.3. *The connection between gravitation and electromagnetism*

On the basis of Theorem I, Hilbert concluded that the four equations (3) are a consequence of the 10 equations (2), such that, '*in the sense indicated* (*in dem bezeichneten Sinne*), *electrodynamic phenomena are effects of gravitation*' (1915a, p. 3; 1915b, p. 397). As this claim is certainly not part of the standard lore of general relativity, it has repeatedly come under severe criticism, most recently by Renn and Stachel (1999, pp. 36–41) and by Corry (2004, pp. 336–337). However, we note that according to what Wheeler termed 'already unified field theory', it has been known for some time that, except for very special conditions of certain null regions, the electromagnetic field is entirely determined by the spacetime geometry, the curvature of spacetime as expressed by the Riemann tensor.[27] Still, since Hilbert relied on a specialized treatment of matter and non-gravitational energy stemming from Mie, we consider only Hilbert's *internal* (to his own theory) justification for this claim.[28] For present purposes, we wish to highlight three results that Hilbert demonstrates, by means of his axiomatic method:

- general invariance, or as we shall prefer to say, general covariance, is connected with the gauge structure of electromagnetism;
- the electromagnetic energy tensor of Hilbert's generally covariant theory yields that of Mie in the special relativistic limit;
- the gravitational equations entail four mutually independent linear combinations of the electromagnetic equations and their first derivatives.

In our opinion, the first and third of these results express one of the two central outcomes reached by Hilbert, by means of the axiomatic method: for *any* theory which seeks to combine generally covariant theories of gravitation and electromagnetism, there follow strong restrictions on the form of the electromagnetic part of the theory as a

(*footnote continued*)
non-standard designation of the electromagnetic potential as well as his practice of using roman letters as indices for that potential and for the 'world parameters'.

[27] Misner & Wheeler (1957), Geroch (1966). These papers follow up earlier results of Rainich. See, e.g., Rainich (1925, p. 498): 'It is often thought that the theory of curved spacetime (general relativity theory) accounts for gravitation but *does not account for the electromagnetic phenomena.* This is not so'.

[28] Hilbert's treatment of energy is discussed in detail in Sauer (1999, pp. 554–557).

consequence of the structure of the gravitational part of the theory.[29] However, we must point out that Hilbert also regarded the second result, concerning the Mie tensor (see immediately below), as a central achievement of his theory, and indeed a promising indication of its general correctness.

The first of the above results is obtained as follows. Hilbert's gravitational equations are expressed as variational derivatives with respect to the metric (Hilbert 1915a, p. 11; 1915b, p. 404):

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0, \tag{4}$$

where the first term is rewritten, in the published version but *not* in the Proofs, with the crucial trace term (see Section 4.1 below),

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right). \tag{5}$$

Now L is a general invariant that is assumed to depend *only* on the $g_{\mu\upsilon}$, the q_s, and the first derivatives $\partial q_s/\partial w^l$. Hilbert had previously shown that from Axiom II (the axiom of general invariance) and a supporting theorem (Theorem II, the Lie derivative of the metric), it follows that L must satisfy the relations (Hilbert, 1915a, p. 10; 1915b, p. 403):

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0. \tag{6}$$

Thus, even though the Mie theory assigns 'absolute' values (a fixed gauge) to the electrodynamic potentials q_s, the matter Lagrangian L in Hilbert's theory depends only on the antisymmetrized derivatives of the q_s

$$M_{ks} = Rot(q_s) \equiv q_{sk} - q_{ks}, \tag{7}$$

that is, on the electromagnetic field tensor. Only by additional assumption is this also the case with Mie's original theory but, of course, that theory is not generally invariant (Born, 1914, p. 28). As Hilbert did not fail to observe, this is a necessary condition for recovering Maxwell's theory. Hilbert has thus shown that the gauge structure of electromagnetism follows from general covariance and the other assumptions for L, summarizing in italic type:

> *This result* [, *on which the character of Maxwell's equations depends*,] *follows here essentially as a consequence of general invariance, hence on the basis of Axiom* II. (Hilbert, 1915a, p. 10; 1915b, p. 403. The bracketed expression does not appear in the Proofs.)

The assumption that nothing else beyond the $g_{\mu\upsilon}$ (but no derivatives of the metric), the q_s, and the so-constrained first derivatives $\partial q_s/\partial w^l$ enter into L, is certainly crucial to this result, and is fully in line with Mie's hypothesis that no matter field quantities of non-electromagnetic nature appear in L (Born, 1914, pp. 24–25). It also has consequences for the interpretation of the energy-momentum tensor $T_{\mu\upsilon}$ in Hilbert's theory. Since all non-gravitational energy/matter is contained in L, it is entirely sufficient

[29]Hilbert's considerations on the tension between general covariance and causality (see Section 4.2 below) are the other central outcome.

for forming $T_{\mu\nu}$, i.e.,

$$\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = \sqrt{g}T_{\mu\nu}. \tag{8}$$

In this respect, Hilbert's gravitational field equations (modulo the considerations of Section 4.1 below), while having the same form as Einstein's

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \chi T_{\mu\nu}, \tag{9}$$

do not have the same interpretation, because Hilbert assumed a particular hypothesis about the electromagnetic constitution of all matter (Earman & Glymour, 1978, p. 303; Sauer, 1999, p. 564). Given this interpretation of $T_{\mu\nu}$, Hilbert is then able to show that the matter tensor of his theory yields the electromagnetic energy tensor of Mie's theory in the special relativistic limit (see Sauer, 1999, p. 555). This is also a fundamental result, and Hilbert here underlined its significance in print with *Sperrdruck* type:

> *Mie's electromagnetic energy tensor is nothing other than the generally invariant tensor obtained by derivation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in the* [special relativistic] *limit*—a circumstance that first indicated to me the necessary close connection between Einstein's general theory of relativity and Mie's electrodynamics, and which convinced me of the correctness of the theory developed here (Hilbert, 1915a, p. 10; 1915b, p. 404).[30]

As Pauli (1921, §55) observed, this is the first demonstration of the now-familiar fact that the energy-momentum tensor of matter (though specialized by Hilbert to Mie's theory) can be obtained by varying the matter Lagrangian with respect to the metric. Moreover, it must be remembered that this 'necessary close connection' between the two theories has been established through the axiomatic method, and so will count toward the triumph of that method as proclaimed by Hilbert at the end of his paper.

Finally, Hilbert demonstrated the connection between the field equations of gravitation and electromagnetism. Using the Lagrangian form of his gravitational equations in conjunction with a version of the contracted Bianchi identities derived in his Theorem III (and which follow from Theorem I), Hilbert arrives at four linearly independent identities containing the Euler derivative associated with the electromagnetic equations,

$$\sum_m \left(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m \right) = 0, \tag{10}$$

where we recall that

$$[\sqrt{g}L]_h = 0 \tag{11}$$

are the abbreviated Lagrangian form of the electromagnetic field equations (Hilbert, 1915a, p. 10; 1915b, p. 406; for discussion, see Sauer, 1999, pp. 556–557). Thus, Hilbert has shown that the gravitational field equations in conjunction with the postulate of general invariance yield four mutually independent combinations of the electromagnetic field equations and their first derivatives. This is the sense in which the electromagnetic phenomena are consequences of the gravitational. Referring back to the assertion that he

[30]Note that *already in the Proofs*, Hilbert referred here to 'Einstein's general theory of relativity' (*der Einsteinschen allgemeinen Relativitätstheorie*), explicitly according due credit to Einstein.

made following his statement of Theorem I, Hilbert immediately claimed, in italic type for emphasis:

> *This is the entire* [*exact*] *mathematical expression of the above generally stated assertion concerning the character of electrodynamics as an accompanying phenomenon* (*Folgeerscheinung*) *of gravitation*. (1915a, p. 12; 1915b, p. 406, with indicated word change in brackets.)

We wish to stress that Hilbert clearly viewed this result, as well as the just-mentioned recovery of Mie's tensor in the special relativistic limit, as *central achievements of his theory. Neither of these has to do with the explicit formulation of the generally covariant field equations of gravitation*.[31] Of course, Hilbert's interpretation of the significance of Theorem I rests on the special choice of H (and L), and the related assumption of the electromagnetic constitution of matter that furnishes the definition of Hilbert's energy-momentum tensor above. However, the use of these assumptions is entirely in line with Hilbert's purpose: that of applying the axiomatic method to the current state of physics. Vizgin, calling attention to the fact that Hilbert's remark that Theorem I was the 'guiding theme (*Leitmotiv*) for the construction of my theory', correctly observes:

> Thus Hilbert's 'Theorem I' a special case of Noether's second theorem, made it possible to regard the equations of electrodynamics as consequences of the gravitational field equations (Vizgin, 1994, pp. 58–59).[32]

The chosen modality 'made it possible to regard' must be emphasized, in light of the assumptions under which Hilbert reached this conclusion.

In essence, although formulated more broadly for any generally invariant theory, Theorem I affirms very clearly, and for the first time, a property of general relativity that is now well known and is indeed associated with Noether's second theorem (Noether, 1918), although this is not always explicitly stated. As recognized in Einstein's canonical 1916 paper on general relativity, by virtue of the requirement of general covariance 'the field equations of gravitation contain four conditions which govern the course of material phenomena'. 'These give', continued Einstein (in this paper, completed in March 1916), 'the equations of material phenomena completely, if the latter is capable of being characterized by four differential equations independent of one another'. In justification of this assertion, Einstein gave an explicit reference to Hilbert's First Communication, and presumably to the page on which appears Theorem I.[33]

Since Einstein was, presumably, aware of Hilbert's declaration that this theorem was the 'guiding motivation' (*Leitmotiv*) of his construction, we can hardly agree with the

[31] Rowe (2001, p. 404) observes that it was 'microphysics not gravitation that Hilbert saw as the central problem area'. We broadly agree that gravitation was not Hilbert's primary focus.

[32] However, in view of the just given derivation, we do not agree with Vizgin's subsequent analysis (pp. 61 *ff)* that Hilbert's choice of just *which* 4 equations followed from the other $n-4$ (according to Theorem I) was arbitrary.

[33] Einstein (1916a), in the Parret and Jeffrey translation, p. 151. In the original text (1916a, p. 810; reprinted 1996, p. 325), Einstein referred to page 3 of Hilbert (1915b); as Sauer (1999, p. 544, note 74) observes, the offprints of this article had pagination beginning with 1. A note in the (1996) text tells us that Einstein's reference should be page 395 (the first page) of the published version. In fact, Theorem I appears on page 397 of that version. See Janssen & Renn (2007) on Einstein's prior use of variational techniques and his recognition that energy conservation is connected with the four identities.

assessment of Rowe (2001) that this sole reference to Hilbert's work in Einstein's canonical exposition of his theory in 1916 is evidence that 'Einstein … could afford to virtually ignore Hilbert's paper…' (p. 412). Nonetheless, the significance of the four identities associated with Noether's second theorem only gradually became established in the literature on general relativity. The Noether identities lead directly to the contracted Bianchi identities, which in the modern view are most often interpreted geometrically (e.g., Trautman, 1962). The Bianchi identities can also be understood as four conditions on the matter-energy-momentum tensor, interpreted as a generalization of the differential energy-momentum conservation laws of matter in the presence of a gravitational field. It was in just this way that the identities associated with Noether's second theorem were interpreted by Pauli in his classic monograph on the theory of relativity (Pauli, 1921, §55).

4. Comparison of the Proofs with the published version of the First Communication

Felix Klein is reported to have commented on the 'completely disordered' character of Hilbert's First Communication, remarking that it was 'evidently a product of great exertion and excitement').[34] We do not know whether Klein was referring to the Proofs or the published version, or perhaps to both, but there is evidence that Klein and Einstein each encountered considerable difficulty in understanding Hilbert's highly formalistic treatment of energy in the published version.[35] In any case, both versions affirm the overall significance of the project as an application of the axiomatic method, and both agree on the main results that follow from this application. As we have already noted, in our opinion, these results are two: (1) showing that for *any* theory which seeks to combine generally covariant theories of gravitation and electromagnetism, there follow strong restrictions on the form of the electromagnetic part of the theory as a consequence of the structure of the gravitational part of the theory (see Section 3.3), and (2) pinpointing a puzzling issue concerning the nature of causality (see Section 4.2) in the new physics of general invariance.

The main differences between the versions are twofold. First, there is the absence in the Proofs of the explicit form of the field equations (see Section 4.1, below), and secondly, there is the absence in the published version of both a clear statement of the problem of causality as well as the solution that appears in the December Proofs (Section 4.2, below). It is the first of these differences that has received considerable attention in the recent literature, and within this there is one point that bears directly on the interpretation of Hilbert's project, which we wish to particularly emphasize. The second difference has been largely neglected, but is—as we shall argue—of crucial importance to the interpretation of the Second Communication.

4.1. Einstein's field equations

The Proofs bear a printer's stamp of 6 December 1915. The published version bears the date of 20 November 1915 as the date of its submission to the *Nachrichten* of the Royal

[34]Klein, Göttingen Lecture Notes, 10 December 1920 (Klein *Nachlaß* XXII C, p. 18), as cited and translated in Rowe (2002, p. 61).

[35]See Klein (1917) for his 'simplification' (*Vereinfachung*) of Hilbert's treatment; for discussion, see Rowe (1999, pp. 212–213) and Brading (2005). For Einstein's difficulties, see the letters of Einstein to Hilbert of 25 and 30 May, and 2 June 1916, and Hilbert to Einstein of 27 May 1916 (Einstein, 1998, pp. 289–295).

Göttingen Academy of Sciences: this is five days *before* Einstein presented the final form of his generally covariant gravitational field equations to the Prussian Academy in Berlin. As finally published on 31 March 1916, Hilbert's First Communication identifies the gravitational part of his 'world function' as the Riemannian curvature scalar density (on which all modern treatments agree) and provides a derivation from a variational principle of what are essentially the same (with the qualifications of Section 3.3 above) generally covariant gravitational field equations as those of Einstein. However, in the Proofs these equations, also based on the Riemann scalar, as well as Hilbert's electrodynamic equations, appear only in their Euler–Lagrange variational form. In contrast, the fully covariant gravitational equations occur explicitly in Einstein's 25 November presentation to the Prussian Academy, appearing in print already on 2 December.

The attention to dates is not mere pedantry for it raises a number of questions, including whether, as several scholars have recently alleged, or insinuated, the Proofs provide evidence of Hilbert's 'nostrification' of Einstein's final results,[36] and even of unethical behavior on Hilbert's part. More important, for our purposes, is the claim in Renn and Stachel's analysis that prior to the publication of Einstein's field equations, Hilbert's research program, as represented in the Proofs, essentially combined Einstein's earlier *non*-covariant *Entwurf* theory of gravitation with Mie's theory.[37] We reject this claim, and are at pains to do so since the requirement of general covariance (or general invariance, in Hilbert's terminology) is utterly fundamental to Hilbert's approach—it is the cornerstone of the epistemological framework within which both his First and Second Communications are formulated. To consider general covariance as 'optional' for Hilbert is to gravely misunderstand and misrepresent his project (see Section 8, below).

In support of our position, and against that of Renn and Stachel, we compare the following aspect of the published version and the Proofs. In the published version, Hilbert identifies his world function as composed of two additive parts,

$$H = K + L, \tag{12}$$

the gravitational and matter components (see note 25 above). In the 1924 republication of this paper, this supposition is stated as an additional Axiom (III) further specifying the world function. However, in the Proofs, some text is missing containing an equation numbered (17). The above specification occurs at the corresponding place in the published version, and *so was almost certainly contained in the Proofs* (Sauer, 2005). Now in both the Proofs and the published version, K is identified, without proof, as the only scalar invariant

[36] In a widely cited letter of 26 November 1915, Einstein complained to his friend Heinrich Zangger in Zurich, that 'only one colleague has actually understood' his new theory and that person had 'sought to "nostrify" (*nostrofizeren*) it ([Max] Abraham's expression) in a clever way'—a clear reference to Hilbert (Einstein, 1998, pp. 204–205). Corry (2004, pp. 99, 419–422), however, observes that this term is 'ambivalent and subtle', and notes that it was 'widely used' to describe the appropriation of existing ideas by Hilbert and his students or collaborators in the axiomatic or mathematical treatment of a given discipline. This seems right to us. Einstein's complaint would seem not so much to be over the 'priority question' but rather with the use Hilbert made of certain of Einstein's ideas (such as treating the metric tensor as gravitational potentials) in tying gravitational theory to Mie's theory of matter. In any case, Einstein and Hilbert were quickly again on good terms; see Einstein's letter to Hilbert of 20 December 1915 (Einstein, 1998, p. 222).

[37] Renn & Stachel (1999, p. 35), Stachel (1999, p. 359). To the contrary, we concur with Sauer (1999, p. 547): 'Hilbert had probably realized that his theory in any case implied field equations which differed from the ones of Einstein's *Entwurf* theory or from those put forward in Einstein's first November communication'.

depending only on the $g_{\mu\nu}$ and its first and second derivatives (Hilbert, 1915a, p. 8; 1915b, p. 402). With some charitable latitude, this permits its identification as the Riemann curvature scalar, $K = g^{\mu\nu}K_{\mu\nu}$, where $K_{\mu\nu}$ is the Ricci tensor.[38] As noted above, the gravitational field equations then appear several pages later in the Proofs as Lagrangian derivatives (Eq. (4)):

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0.$$

As the derivation begins with a generally invariant world function, and since Lagrangian differentiation with respect to the metric is a covariant operation, Hilbert's gravitational equations (26) are generally covariant not only in the published version *but also already in the Proofs*.[39]

In the published version, but *not* in the Proofs, Hilbert noted that 'it follows easily without calculation' (1915b, 405) that (Eq. (5)):

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right).$$

The appearance of the explicit form of the field equations in the published version, including the crucial 'trace' term, occurs in conjunction with Hilbert's removal of the intricate non-covariant energy theorem that he constructed in the Proofs (see Section 4.2, below). This latter change indicates Hilbert's realization that no restriction on the spacetime coordinates is required for energy-momentum conservation, and indeed Einstein pointed out in his 25 November paper that energy-momentum conservation is a consequence of his generally covariant field equations. Since so much of the text of the Proofs was devoted to constructing the non-covariant energy theorem, Corry, Renn, and Stachel (1997) allege that 'knowledge of Einstein's result may have been crucial to Hilbert's introduction of the trace term into his field equations' (p. 1272). However, we think this—and the implications that Corry, Renn, and Stachel draw from it—must be taken with a grain of salt for, on the one hand, Sauer has shown that the calculation of the Einstein tensor (containing the trace term) follows rather naturally from Hilbert's assumptions and his Theorem III, which essentially recovers the contracted Bianchi identities.[40] In fact, Hilbert presented an explicit calculation only in the edited 1924 republication of the papers (1915b) and (1917) in the *Mathematische Annalen.*[41] On the other hand, viewing Hilbert's axiomatic construction as a whole, we do not believe that an explicit evaluation of the gravitational field equations in tensor form was a particularly important goal of that

[38]Hilbert failed to state that the identification as the Riemann scalar requires that K contains the second derivatives of $g_{\mu\nu}$ only linearly, for reasons presumably known to Hilbert but clearly amplified by Landau & Lifshitz (1975, p. 268). Rowe (2001, pp. 417–418) argues that Hilbert relied on a bit of local Göttingen 'mathematical folklore' regarding differential invariants. An explicit proof, but under the assumption of the positive definiteness of the metric, is given in Weyl (1921, Appendix 2; English trans., pp. 315–317). Both Rowe and Weyl credit the result to a paper by Felix Klein's assistant Hermann Vermeil (1917). On the latter's contribution, see Sauer (2005).

[39]This has been particularly emphasized by Sauer (1999, p. 547).

[40]Sauer (1999, p. 564): 'The argument may not follow so easily without calculation but is nonetheless true if it is understood that the second derivative of the metric tensor enters only linearly and if the condition is taken into account that the combination of $K_{\mu\nu}$ and $g^{\mu\nu}K$ has to satisfy the contracted Bianchi identity … derived in Hilbert's Theorem III'.

[41]For discussion of editorial changes in the 1924 republication, see Renn & Stachel (1999, pp. 64–65).

project. Indeed, we think that the weight of evidence agrees with Corry's (2004, p. 403) assessment that

> Hilbert arrived at general relativity in a roundabout way, while pursuing a much more general aim: a unified, axiomatic foundation for all of physics.

So even though the *explicit* generally covariant form of the field equations does not appear in the Proofs, nevertheless—as already noted—Hilbert's gravitational field equations as given *implicitly* there in terms of his Lagrangian *are* generally covariant.

4.2. Hilbert's target: the 'problem of causality'

Both the Proofs and the published version of the First Communication contain Hilbert's Axiom I (the 'world function', containing the gravitational and electromagnetic potentials), his Axiom II (of general covariance, or 'general invariance' in Hilbert's terminology), and his Theorem I. In both versions Hilbert declares that Theorem I is the '*Leitmotiv*' of his theory, thereby indicating that it is his principal concern. In the Proofs, but not in the published version, Hilbert explicitly spells out the implications of Theorem I for his system of fundamental equations of physics (1915a, pp. 3–4):

> Our mathematical theorem teaches that the above axioms I and II can yield for the 14 potentials only 10 equations essentially independent of one another. On the other hand, by upholding general invariance, no more than 10 essentially independent equations for the 14 potentials $g_{\mu\upsilon,}$ q_s, are possible at all. Therefore, if we want to preserve the determinate character of the fundamental equations of physics according to Cauchy's theory of differential equations, the requirement of four additional non-invariant equations supplementing (2) and (3) is essential.

Thus, independent of the physical validity of his system of fundamental equations, for which he adduced no evidence whatsoever, Hilbert clearly underscored his interest in the fact that the mathematical underdetermination in question (10 independent equations for 14 potentials) is solely a consequence of his axiom of general invariance as applied to the potentials stated in Axiom I.

As befits its preeminent concern with the *consistency* of all axioms and assumptions undergirding a theory, the axiomatic method has revealed an apparent tension between general covariance and causality in the sense of a failure of univocal determination, a conflict characterized in terms of whether *any* theory satisfying Axioms I and II admits of a well-posed Cauchy problem.[42] Theorem I suggests that it is a property of any such theory that it does not.[43] The Cauchy problem, for a system of second-order partial differential equations, is to show that from given initial data assignments to the unknown field functions and their first (time) derivatives in a bounded region, the initial data yield unique solutions to these equations as far as possible from that region (the region's 'domain of dependence'). For field theories formulated in spacetime, the initial data are formulated on a given spacelike hypersurface Σ, and the essential problem is that of showing that the field equations determine the second time derivatives of the given field quantities. As Hilbert repeatedly emphasized, in all physics prior to general relativity (i.e., in all prior theories

[42] See Appendix A for a sketch of the Cauchy problem in general relativity.

[43] We discuss Hilbert's analysis in relation to Einstein's 'hole argument' in Section 7 below.

admitting a variational formulation), Cauchy determination required that there be precisely as many independent equations as there are independent functions to be determined. However, the situation is complicated in a generally covariant spacetime theory by the freedom to make arbitrary coordinate transformations (equivalently, diffeomorphic point transformations) of solutions to the field equations. As stipulated for a generally invariant Lagrangian by Hilbert's Theorem I, this is the fact that not all the Euler–Lagrange equations obtained by variation of the integral invariant with respect to the field quantities and their derivatives are independent. More precisely, 4 of these are always the result of the remaining $n-4$ spacetime equations. Thus, Theorem I is *a precise mathematical statement of the tension between the postulate of general covariance and the requirement of causality in the mathematical sense of univocal determination.*

Notice that univocal causal determination—in the sense required by a well-posed Cauchy problem—is not an axiom in Hilbert's construction. Nevertheless, it is a requirement satisfied by all previous field theories, and so its seeming failure in the context of general invariance surely sparked Hilbert's interest, a topic to which we turn in Section 7. But as we have repeatedly stated, in our opinion this is one of the two central outcomes that Hilbert reached by means of the axiomatic method: *any generally covariant theory raises deep questions about causality, in both the mathematical and (as we shall see) the physical sense.*

Hilbert's diagnosis in turn marked out a strategy for resolving the apparent tension between general covariance and failure of univocal determination: to find, if possible, four equations additional to the 10 independent equations that will render the Cauchy problem well posed. Finding the 'four additional non-invariant equations' is the motivation behind the intricate mathematical construction in the Proofs of an 'energy form':

$$E=\sum_{s} e_s p^s+\sum_{s,l} e_s^l p_l^s. \tag{13}$$

Here e_s is termed the 'energy vector', and p^s is an arbitrary contravariant vector. We have used Hilbert's notation: there is no summation convention in use, and the subscript indices indicate coordinate derivatives. The 'energy form' is constructed from the tensor density $\sqrt{g}P_gH$, where P_g is a differential operator on the world function H. A prime consideration both here, and in the different treatment of energy in the published version, will be to recover Mie's energy tensor as a special case (see below). Hilbert found four supplementary equations by re-writing his 'energy form' to include an expression whose vanishing would correspond 'to the energy theorem of the old theory,

$$\sum_{i} \frac{\partial e_s^l}{\partial w_l}=0, \tag{14}$$

and then requiring that, *for special spacetime coordinates* w_k *adapted to this 'energy theorem'*, the theorem holds.[44] Accordingly, the 'energy theorem' is not generally

[44]We rewrite the energy form as $E=\sum_s e_s p^s+\sum_{s,l}(\partial/\partial w_l)(e_s^l p^s)+\sum_{s,l} p^s(\partial e_s^l/\partial w_l)$, and discard the divergence term to arrive at $E=\sum_s e_s p^s+\sum_{s,l} p^s(\partial e_s^l/\partial w_l)$. Hilbert then notes that the 'energy theorem' holds iff $e_s=0$, which in turn holds iff $(d^{(g)}\sqrt{g}H/\mathrm{d}w_s)=0$. Notice that it is the *coordinate* derivative of $\sqrt{g}H$ that vanishes when the 'energy theorem' holds, not the *covariant* derivative.

covariant, and Hilbert used it to supplement the generally covariant field equations, as stated in a third, and final, axiom appearing only in the Proofs:

Axiom III. ('The Axiom of Space and Time'): 'The space-time coordinates are such particular world parameters for which the energy theorem (14) is valid'.

Elucidating this result, Hilbert clarified the main point, that these four non-covariant equations complete the system of fundamental equations of physics (1915a, p. 7):

> On account of the same number of equations and of definite potentials, the causality principle for physical happenings (*Geschehen*) is also ensured, and with it is unveiled to us the narrowest connection between the energy theorem and the principle of causality, in that each conditions the other.

The idea that satisfaction of energy conservation (the energy theorem (14)) requires four non-covariant equations is almost certainly taken from the Einstein and Grossmann *Entwurf* theory (1913),[45] where four non-generally covariant equations ensure energy conservation by restricting the covariance class of the field equations. But Hilbert's rather more complicated construction has, philosophically and motivationally, a different *raison d'être*. We thus reject the view of Renn and Stachel (1999, p. 73) who regard Hilbert's energy construction, intended to restore causality, as his 'Proofs argument against general covariance'. Rather, Hilbert's four non-generally covariant equations ensuring energy conservation are used to *extract a Cauchy-determinate structure within an otherwise generally covariant theory* (and *not* to abandon general covariance).[46,47] We return to this in our discussion of the 'hole argument', below (Section 7).

As it happened, the very complex mathematical derivation in the Proofs leading to Hilbert's four energy equations was cut, together with *all* of its motivation, from the published version. The reason is that, in the light of Einstein's 25 November presentation of his field equations to the Berlin Academy (Einstein, 1915), this turned out to be the wrong approach for solving the tension between general covariance and Cauchy-determination. Hilbert dropped it altogether, significantly modifying and truncating his treatment of energy. There—consistent with the implicitly generally covariant energy in Einstein's treatment of 25 November—Hilbert derived a generally covariant 'energy equation' which anyway is consonant with the 'trace' term in the gravitational field equations popping out through explicit calculation from their Lagrangian derivatives.

[45] Renn & Stachel (1999, p. 32) report Einstein's conviction '[e]ven before Einstein developed the hole argument', that energy-momentum conservation requires such a restriction.

[46] This is also pointed out by Sauer (2005, n. 5): 'Hilbert kept the generally covariant field equations as fundamental field equations and only postulated a limitation of the physically admissible coordinate systems'. Yet Sauer does not make enough of this, we think. Earlier in his text he writes that Hilbert's Axiom III is a *restriction* of the general covariance of Hilbert's theory, there seeming to subscribe to the view that Hilbert followed Einstein in seeking to limit the covariance of his theory.

[47] As Sauer (2005, n. 5) observes, Janssen & Renn (2007) reserve the terminology 'coordinate restrictions' to apply to Einstein's use of energy conservation whereby the covariance properties of the fundamental field equations themselves are restricted. Hilbert's use differs significantly, in implying no such restriction on the covariance properties of the field equations, and we therefore use the terminology of 'coordinate conditions' in our discussion of Hilbert. Thus, *Hilbert used the four energy equations stated in his energy theorem (15) to impose coordinate conditions on the generally covariant field equations*. However, as Sauer also notes, Hilbert's conditions in the Proofs differ from the modern understanding of coordinate conditions since *all applications* of the field equations require their satisfaction.

Nevertheless, the issue of causality in a generally covariant theory doesn't go away for Hilbert. We claim that the Second Communication contains his much revised, and lengthy, reconsideration of this issue, and that the entirety of this paper is rightly understood only in this light.[48]

5. The First Communication and the axiomatic method

We recall that the task of the axiomatization of physics was the sixth in the famous list of 23 mathematical problems Hilbert posed at the International Congress of Mathematicians in Paris in 1900.

> The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part .... If geometry is to serve as a model for the treatment of physical axioms, we shall try first by a small number of axioms, to include as large a class as possible of physical phenomena, and then by adjoining new axioms to arrive gradually at the more special theories. ... As he has in geometry, the mathematician will not merely have to take account of those theories coming near to reality (*Wirklichkeit*), but also of all logically possible theories. He must be always alert to obtain a complete survey of all conclusions derivable from the system of axioms assumed. Further, the mathematician has the duty to test in each instance whether the new axioms are compatible with the previous ones. The physicist, as his theories develop, often finds himself forced by the results of his experiments to make new hypotheses, while he depends, with respect to the compatibility of the new hypotheses with the old axioms, solely upon these experiments or upon a certain physical intuition, a practice which is not admissible in the rigorously logical building up of a theory. The desired proof of the compatibility of all assumptions seems to me also of importance, because the effort to obtain such a proof always forces us most effectively toward an exact formulation of the axioms (Hilbert, 1901; English trans. Gray, 2000, pp. 257–258).

Inclusion of the axiomatization of physics among the other purely mathematical problems on his list appears rather incongruous until Hilbert's lifelong interest in physics is taken into account.[49] For our purposes, there are three items of interest in this passage.

- As noted in Section 2, geometry is regarded as a model for the axiomatization of physical theories.
- In axiomatizing, the mathematician is to take account of 'all logically possible theories', not just phenomenological theories 'near to reality', and so the axiomatic method is ideally suited for setting up a speculative theory from whose common basis both gravitational and matter fields might arise.

[48] The topic of energy-momentum in general relativity did not go away: it was the subject of ongoing discussions between Hilbert, Einstein, and Klein (see Brading, 2005), and remains a delicate issue (for discussion, see Hoefer, 2000).

[49] Corry (2004) amply demonstrates the extent of this interest, examining in considerable detail Hilbert's many lecture courses and seminars devoted to various physical theories or questions of current physics.

- Axiomatization has the express purpose of testing the consistency of new hypotheses with previously adopted axioms and assumptions, a task that requires 'the rigorously logical building up of a theory' in place of its informal statement in experiential or intuitive terms.

These points are of special interest for understanding the role of the axiomatic method in Hilbert's two notes on the 'Foundations of Physics'; in particular, they highlight again Theorem I's epistemological significance, pinpointing the tension between the apparently conflicting assumptions of general covariance and causality. Above all, we wish to stress in general the *hypothetical* character of Hilbert's axiomatic approach to physics, which was explicitly recognized by Hilbert's former student and Göttingen physics colleague Max Born in a tribute on the occasion of Hilbert's 60th birthday entitled 'Hilbert and Physics':

> [B]eing conscious of the infinite complexity he faces in every experiment [the physicist] refuses to consider any theory as final. Therefore … he abhors the word 'axiom' to which the sense of final truth clings in the customary mode of speech. … Yet the mathematician does not deal with the factual happenings, but with logical connections; and in *Hilbert's* language the axiomatic treatment of a discipline in no way signifies the final setting up of certain axioms as eternal truths, but the methodological requirement: Place your assumptions at the beginning of your considerations, stick to them and investigate whether these assumptions are not partially superfluous or even mutually inconsistent (Born, 1922, pp. 90–91).

As both Einstein and Hilbert were aware in 1915, Einstein's gravitational theory, though in principle capable of encompassing all matter fields into spacetime geometry, did not itself suppose any particular theory of matter. This can be most readily seen in the purely phenomenological significance it accords to the stress-energy tensor, as a place holder into which any detailed theory of matter must fit or conform, a mere structure of 'low grade wood' contrasting with the 'fine marble' of the left-hand (geometric) side of the Einstein field equations (Einstein, 1936, p. 335). On the other hand, the axiomatic method seems ideally suited for setting up a speculative theory from whose common basis both gravitational and matter fields might arise. In this way, a system of fundamental equations of physics might be erected that would include all known physical interactions. The principal attraction of the Mie theory to Hilbert seems precisely to have been that, coupled with Einstein's theory of gravitation, it enabled such a hypothetical axiomatic completion of physics that could be studied by drawing consequences from the amalgamation of the two theories. In this regard, Hilbert's own 'theory' of 1915 is a canonical illustration of a mode of investigation by the 'axiomatic method', in Hilbert's most precise characterization of that method, as the 'mapping' (*Abbildung*) of a 'domain of knowledge' (*Wissensgebiet*) onto

> a framework of concepts so that it happens that the objects of the field of knowledge correspond to the concepts, and the assertions regarding the objects to the logical relations between the concepts. Through this mapping, the (logical) investigation becomes entirely detached from concrete reality (*Wirklickkeit*). The theory has nothing more to do with real objects (*realen Objekten*) or with the intuitive content of knowledge. It becomes a pure construction of thought (*reine Gedankengebilde*), of which one can no longer say that it is true or false. Nevertheless, this framework of

> concepts has a significance for knowledge of reality in that it presents a possible form of actual connections. The task of mathematics is then to develop this framework of concepts in a logical way, regardless of whether one was led to it by experience or by systematic speculation.[50]

Hilbert was familiar with the Mie theory at least since its discussion in the Göttingen Mathematical Society in December 1912 and again in December 1913, when Born had set it into a more canonical mathematical form (Corry, 1999, p. 176). Certainly, that Mie had sought to derive field equations of a generalized Maxwellian electrodynamics from an axiom of a Lorentz (orthogonally) invariant 'world function' appearing as a variational principle, fitted very naturally into Hilbert's axiomatic approach.

But the Mie theory presented an eminently suitable candidate for the attention of the 'axiomatic method' for a number of other reasons that merit illumination, reasons not so much physical but mathematical and philosophical. In particular, Hilbert saw distinct advantages in the Mie theory over the only other rival electromagnetic theory of matter of consequence in 1915, the electron theory, on which Hilbert had lectured in Göttingen in the summer semester of 1913, and would again in the summer semester of 1917.[51] Namely, the Mie theory was *a priori* consistent with *the principle of causality* in two ways that the electron theory was not.

First, it employed only differential equations, whereas the electron theory, as Hilbert noted in lectures in the summer semester of 1916 (pp. 101–102), was a mixture (*ein Gemisch*) of functional, differential, and integral equations. From the standpoint of consistency with the field theoretic prohibition against action-at-a-distance, the Mie theory was clearly to be preferred to the electron theory.

Second, the Mie world function yielded four electrodynamical equations for the four unknown electrodynamic potentials. From given boundary and initial conditions, one could show that the state of the world at any future time could be univocally determined via these equations through specification of the values of these potentials at any prior time, just what is required by the principle of causality (as Hilbert understood that principle). Although in the Mie theory itself this causal determination is purchased at the cost of gauge invariance (the Mie potentials have 'absolute' values), we have seen that Hilbert's construction recovers the gauge invariance of electromagnetism.[52] Hence, causality comprised another prop of support to the Mie theory. Ironically, precisely what current wisdom deems wrong about the Mie theory, that it assigns absolute values to the electromagnetic vector potentials, was thus a philosophical ground in favor of it cited by Hilbert. In sum, even as Hilbert was still uncertain about the standing of the principle of causality in the new physics of Einstein's principle of general invariance (e.g., Hilbert 1916a, p. 110), he found Mie's theory most suitable for

[50]Hilbert's Winter Semester 1921–1922 lectures on the '*Grundlagen der Mathematik*', cited and translated in Hallett (1994, pp. 167–168).

[51]Corry (1999, pp. 174, 183). Corry (2004, p. 271) observes that 'Hilbert's lectures on electron theory emphasized throughout the importance of Lorentz transformations and of Lorentz covariance, and continually referred back to the works of Minkowski and Born'.

[52]As a referee has reminded us, within the broad framework of Mie's theory, conceivably one might hope to find a matter representation based on generalized Maxwell equations following from a Lagrangian containing only gauge invariant terms.

incorporation into his theory by its *a priori* consistency with the requirement of causality.

Finally, as we saw in Section 3.3, above, there were also *a posteriori* reasons justifying Hilbert's incorporation of Mie's theory into his axiomatization of fundamental physics. Namely, Hilbert could show that the gauge structure of electromagnetism was recovered by his generally covariant generalization of Mie's theory, and that his energy tensor for non-gravitational energy coincided with Mie's energy tensor in the special relativistic limit. As we have seen, both of these results were crucial to Hilbert's claim that electrodynamic phenomena are a consequence of gravitation.

6. Hilbert's Second Communication on 'The Foundation of Physics'

On 4 December 1915, 'The Foundation of Physics (Second Communication)' was presented to the Göttingen Academy. A further presentation under that rubric was delivered on 26 February 1916. However, both of these were withdrawn before publication, and no version of either has apparently survived (Sauer, 1999, p. 557 and n. 120, p. 560 and n. 129; 2001, p. 3). From Hilbert's correspondence, it may be conjectured that one principal topic concerned a much anticipated derivation of 'the electron' from Hilbert's generalization of Mie's electrodynamic equations. In a brief summary, Sauer emphasizes that Hilbert could not provide what had been promised at the conclusion of the First Communication, viz., the electromagnetic part of a Lagrangian world function that would allow derivation of the electron, and thus explain (as Hilbert put it) 'the most intimate and up to now hidden processes within the atom'. That goal, certainly, was never attained. A third submission under that title was finally given to the Academy for publication on 23 December 1916, appearing in print early in 1917. According to our reconstruction, this third submission must have a significantly different content from the previous (and non-extant) versions. Despite period recognition by Weyl, Pauli, and von Laue, the paper has been treated as largely independent of Hilbert's First Communication, and in any case, has been widely disregarded by historians, physicists, and philosophers alike. We claim that the Proofs enable us to see that the two communications are united within a logical and epistemological investigation of a generally covariant field physics via the axiomatic method.

As discussed above, in the light of Einstein's final submission of 25 November, Hilbert's resolution of the twin-faceted problem of causality in the Proofs could not be maintained. Hilbert had made use of a non-generally covariant treatment of energy to resolve the problem of mathematical underdetermination, whereas Einstein had demonstrated that the general theory of relativity admits a generally covariant treatment of energy. It is our contention that the primary concern of the published version of the Second Communication is the presentation of a new resolution of the causality problem. This new resolution required a richer and deeper physical and epistemological approach, which amounts to a revision of Kant in the light of the new physics of general covariance. We discuss this claim in detail in Section 8, below, following our presentation of the details of the Second Communication. However, it is necessary for this presentation that we indicate the pertinent features of Hilbert's revision of Kant. We do this in Section 6.2, before outlining Hilbert's paper in Sections 6.3–6.7. We begin, however, with a brief glance at the scant secondary literature on the Second Communication.

6.1. Secondary literature on the Second Communication

There is remarkably little modern discussion of Hilbert's Second Communication, even in the history of relativity literature. Pais's assessment (1983, p. 258) that 'it contains a synopsis of [Hilbert's] 1915 paper and a sequel to it' is perhaps typical. Such neglect did not occur in the wake of its appearance. Weyl, remarkably, had originally cited this paper, but not the First Communication, in the penultimate version of the first edition of his classic *Raum-Zeit-Materie* (1918a), much to Hilbert's chagrin when he read the printer's proofs.[53] Pauli, who needed Klein's prodding to credit Hilbert (1915b) with simultaneous discovery of the Einstein gravitational field equations, nonetheless devoted a section of his relativity book with the arresting title 'Reality Relations' to Hilbert's Second Communication (see further below).[54] Among recent discussions, Stachel (1993) considers this paper solely as a first preliminary attempt to put the generally covariant field equations of gravitation in Cauchy normal form, from which stems subsequent work on the initial value and Cauchy problems in general relativity.

In our opinion, the Proofs of the First Communication provide a crucial insight into the intended main theme of the Second Communication, presenting the basis for a significant reinterpretation of that paper. It might seem curious that such a reconsideration has not been given before. However, the exclusive interest heretofore in reconsidering the First Communication in the light of the Proofs has been the 'priority issue', concerning whether Einstein or Hilbert first arrived at the Einstein field equations. Given its later date, the Second Communication has not been thought relevant. However, we think that a correct interpretation of the Second Communication, and thereby of the common project carried out by Hilbert jointly in his First and Second Communications, reveals that the attention long given the question of priority is in fact misplaced, stemming from a failure to view Hilbert's project within the context of the axiomatic method within which it was conceived and carried out. Once this context is taken into account, we believe that Felix Klein in 1921 got it exactly right: 'there can be no talk of a question of priority' (see n. 7 above).

By far the most sustained examination we know of Hilbert's second paper is that of Renn and Stachel (1999, pp. 77–90). Their principal conclusion is that the paper shows 'Hilbert at work on general relativity', his choice of topics exhibiting how his 'original goal of developing a unified gravito-electromagnetic theory…has been modified in the light of the successes of Einstein's purely gravitational program' (p. 77). To be sure, much of Hilbert's paper is concerned with integrating his field equations in regions where the electromagnetic field disappears, i.e., where they coincide with Einstein's 'empty space' equations. But according to Renn and Stachel, Hilbert's paper simply tackles *seriatim* a list of six topics *within* general relativity, with no tissue or thread of argument connecting them. Their discussion of the last two of these ('Euclidean geometry' and 'The Schwarzschild Solution'), occupying 11 pages, is however not so much concerned with Hilbert's second paper as with material from Hilbert's 1916/1917 winter semester lectures

[53] See the draft letter of Hilbert to Weyl dated 22 April 1918, as translated in Rowe (2003, p. 66).

[54] Pauli (1921, §22). Rowe (2001, p. 408) notes that Pauli's in-print recognition of Hilbert as codiscoverer came only after two letters from Felix Klein (8 March 1921 and 8 May 1921), the latter complaining that the 'physicists mostly pass over Hilbert's contributions in stony silence'. See Pauli (1979, pp. 27, 31).

where Renn and Stachel find Hilbert's motivations far more transparent. Their conclusion is illustrative:

> In summary, this paper must be considered a curious hybrid between the blossoming of a rich mathematical tradition that Hilbert brings to bear on the problems of general relativity, and the agony of facing the collapse of his own research program (p. 90).

In this section we shall show that Hilbert's second paper is not at all a laundry list of special topics in general relativity and that Hilbert, far from being in 'agony' over the 'collapse of his own research program', deemed this paper to be its completion.

Our contention is that, in the Second Communication, Hilbert is returning to the puzzling issue of Cauchy determination for generally covariant theories clearly articulated in the Proofs of his First Communication. Hilbert's Second Communication is, we argue, principally concerned with providing a satisfactory reconciliation between the principles of general invariance and causality. Hence, the structure of the argument in the second paper must be understood in the light of the suppressed proofs.

6.2. *Outline of Hilbert's revision of Kant*

As we noted in Section 2 above, Hilbert's discussion of generally covariant physics should be understood against the background of his axiomatic method and his appeal to Kantian epistemology. According to Hilbert (and many others, see Ryckman, 2005), the new generally covariant physics requires some modification of Kant. We discuss this in detail in Section 8, below, but it is worthwhile indicating the main features of Hilbert's particular revision now, before turning to our discussion of the details of the Second Communication.

The crux of Hilbert's suggested amendment, in our reconstruction, is that a distinction must be made between being a *possible object of physics* and being a *possible object of experience*. Such a distinction is, of course, anathema to the usual understanding of Kant's conception of cognition, which does not extend beyond objects of possible experience. However, for Hilbert, while representation in space and time and the requirement that causes precede their effects remain conditions of a possible object of experience, they are no longer necessary conditions for a possible object of physics. Rather the guiding criterion of physical objectivity is now general invariance that, of course, Hilbert has adopted as an axiom of his theory. Thus, according to Hilbert, general invariance is an ideal constraint informing the search for fundamental physical laws. As we shall discuss in Section 8, Hilbert viewed this amendment as a significant step forward in eradicating unnecessarily limiting aspects of human subjectivity from the conceptual structure of physics.

In short, the principle of causality (that causes always precede their effects, in distinction from the mathematical requirement of univocal determination) is an *anthropomorphic* condition imposed by the structure of the human mind fashioning the character of human experience, *but this limitation no longer bounds the possible objects of physics*. In Hilbert's view, this is the fundamental philosophical significance of Einstein's theory and it is of vital importance for understanding Hilbert's discussion of causality in his Second Communication, as we will now show.

6.3. Outline of Hilbert's Second Communication

Hilbert begins with several 'preliminaries', in which he introduces the geometrical notions that he will rely on, and then announces his intent (see Section 6.4). The main theme of the paper concerns the relationship between general covariance and the principle of causality, which Hilbert separates into two problems: the problem of causal ordering (see Section 6.5), and the problem of univocal determination (Section 6.6). With the results of this investigation in hand, Hilbert then revisits the 'old question' of whether Euclidean geometry is valid for reality (see Section 6.7).

As we shall see, Hilbert's treatment of the problem of causality in generally covariant theories has four principal facets. First, he observed that arbitrary point transformations (diffeomorphisms) do not respect the relation of cause and effect among world points lying on the same timelike curve. To rectify this, he introduced the notion of '*proper coordinate systems*', transformations among which always respect the distinction between spacelike and timelike coordinate axes and can never reverse the temporal order of cause and effect. Next, he pointed to the consequent need to reformulate the principle of causality within what he termed the 'new physics' of general invariance, showing that here the univocal determination of future states from present states requires coordinate restrictions on the initial data in order to locally describe dynamical evolution off that surface. This is attained by employing a 'Gaussian' coordinate system, a particular type of *proper coordinate system*. The purchase of univocal determination in the new physics at the cost of adopting special coordinate systems prompted Hilbert, thirdly, to state a 'sharper conception' of the principle of general relativity (general invariance) underlying this physics. By means of this sharper conception, he is able to give a clear account of under what conditions a statement of physics is to be regarded as physically meaningful. Finally, Hilbert took up the related issue of the inconsistency of Euclidean geometry (permitting, on account of its globally fixed metrical structure, the concept of action-at-a-distance) with the new physics of fields, which he calls a four-dimensional pseudo-geometry. To this end, he discussed the conditions under which a pseudo-Euclidean (Minkowski) metric arises in the new physics, and he re-derived the external Schwarzschild solution corresponding to the solar gravitational field without the assumption that the $g_{\mu\nu}$ had pseudo-Euclidean values at infinity, that is, that the solar system is embedded in a pseudo-Euclidean world.

6.4. Preliminaries

Hilbert begins with the introduction of a four-dimensional pseudo-geometry. The geometry is 'pseudo', Hilbert explains,[55] because in coordinatizing the space, one coordinate will be distinct from the other three (the metric is *indefinite* rather than positive definite). By taking the coordinates x_s to be functions of parameters along the curves of this geometry, the curves are partitioned in the familiar way into three classes—spacelike (*Strecke*), timelike (*Zeitlinie*), and null (*Nullinie*). Physical significance is accorded to *Strecke* and *Zeitlinie* as follows. Measurements of intervals along *Strecke* by means of 'measure threads' give lengths, and measurements of intervals along *Zeitlinie* by means of 'light clocks' give proper times.

[55]The term 'pseudo-geometry', apparently derived from the term 'pseudo-Euclidean' applied to the metric of Minkowski spacetime, appears to have been local Göttingen mathematical vernacular.

Three aspects of Hilbert's treatment of these now familiar topics merit attention. First, such a pseudo-geometry must allow for mixed orthogonal transformations corresponding to reversals of direction in time. Second, Hilbert's characterization has a purely geometrical flavor. Thus, in the context of his pseudo-geometry, Hilbert wrote down the Monge ordinary differential equation as $g_{\mu\nu}$ $(\mathrm{d}x_\mu/\mathrm{d}p)(\mathrm{d}x_\nu/\mathrm{d}p) = 0$, and the Hamilton–Jacobi partial differential equation as $g^{\mu\nu}$ $(\partial f/\partial x_\mu)(\partial f/\partial x_\nu) = 0$, i.e., in terms of the metric tensor, showing that these are equations of the null cone field, and that geodesic null lines are the characteristics of the Monge, and bi-characteristics of the Hamilton–Jacobi equation. Null geodesics emanating from any world point form the null cone ('conoid') at that point and Hilbert observed that the equation of this conoid is a solution to the Hamilton–Jacobi equation, whereas all timelike worldlines emanating from a world point lie inside its conoid, their boundary (Klein, 1927, p. 107; Renn & Stachel, 1999, p. 79).

Thirdly, it is significant that in his discussion of 'measure threads' and 'light clocks' Hilbert does not, as Einstein will several years later (Einstein, 1921), posit 'practically rigid rods' and 'ideal clocks' as presuppositions of the applicability of the pseudo-geometry to the actual world.[56] As with all instruments of measurement, these measure threads and light clocks presuppose the space and time of human sensibility. And, as with Einstein's rods and clocks, they are independent assumptions added to bring empirical content to the theory. Nevertheless, Hilbert's ideal instruments are less egregiously independent: they lack the pseudo-Euclidean assumptions of Minkowski geometry and, in principle, can be completely characterized within 'pure field theory' (as set up in the fundamental field equations (2) and (3) of his First Communication). To highlight this character, Hilbert spoke of measure threads and light clocks as 'instruments of invariant character' in whose terms 'every physical fact must ultimately (in principle) be ascertained' (1917, p. 61). Presumably, Hilbert meant by this that they can be employed *everywhere* and *everywhen* by any observer in an arbitrary state of motion, and so provide the empirical basis for a generally invariant field physics, of which, as already noted, these instruments are themselves a part.

All of the above has been preliminary. At this point the main theme of the paper appears: the problem of causality for the new physics of general invariance. The problem is twofold. First, there is the question of how to ensure that the new physics respects the experienced causal ordering of events. The second, taken up in Section 6.6 below, revisits the question of the Cauchy problem in the new physics (i.e., the problem of univocal determination), since the resolution given in his First Communication has turned out to be incorrect.

6.5. The problem of causal order

On the basis of Axiom II of his First Communication (the axiom of general invariance), and with implicit reference to Einstein's requirement of general covariance for the fundamental equations of the new physics, all coordinate systems arising from x_s by arbitrary smooth transformations have up to now been regarded as on an equal footing with one another. However, Hilbert observed that a conflict with the causal order will arise if two world *points* lying along the same timelike curve, and standing in the relation of

[56] Torretti (1978, p. 391, n. 8) rightly observes: 'Riemann's physical geometry requires ideally thin, perfectly flexible and inextensible strings…'.

cause and effect, can be transformed so that they become *simultaneous* (i.e., lie on the same data hypersurface). The causal order concerns our *experience* of the world in space and time, and thus we have an apparent conflict between the overriding demand of objectivity expressed by general covariance and the *experienced causal ordering* of events. The conflict is only apparent, however.

Although Hilbert speaks (p. 57) about the need to restrict the arbitrariness *of coordinate systems*, his example concerns *point transformations* (in fact, along one and the same timelike curve) and the fact that the diffeomorphism invariance of manifold point transformations need not preserve the relation of causal order among events. If the new physics is to be compatible with the experienced causal ordering of events, we need to restrict the allowed coordinate systems such that under a coordinate transformation any timelike curve remains a timelike curve. To achieve this end, Hilbert introduced what he called 'proper' coordinate systems.

If x_4 is designated as the 'proper' time coordinate, a 'proper (*eigentlich*) coordinate system' may be defined as one in which the following four inequalities are satisfied by the components of the metric tensor:

$$g_{11}>0, \quad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix}>0, \quad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix}>0, \quad g_{44}<0. \tag{15}$$

These inequalities are justly named 'Reality Relations' ('*Realitätsverhältniße*') by Pauli (1921, §22), for they implement, in the case of general Riemannian geometry, the physical requirement of metrical indefiniteness: that three of the coordinate axes are spacelike, and one timelike. Together the restrictions imply that $g(=\det|g_{\mu\nu}|)<0$, so $\sqrt{-g}$ must replace $\sqrt{g}$ in all tensor formulae. A coordinate transformation carrying such a proper spacetime description into another proper spacetime description is called a *proper* spacetime coordinate transformation.

The desired effect of proper coordinate systems is seen in considering a parameterized timelike curve, $x_s = x_s(p)$. Since for such a curve, and using Hilbert's abbreviation (1917, p. 54),

$$G\left(\frac{\mathrm{d}x_s}{\mathrm{d}p}\right)<0, \quad \left[G\left(\frac{\mathrm{d}x_s}{\mathrm{d}p}\right) = g_{\mu\nu}\frac{\mathrm{d}x_\mu}{\mathrm{d}p}\frac{\mathrm{d}x_\nu}{\mathrm{d}p} \quad (\mu,\nu=1,2,3)\right], \tag{16}$$

it follows that in a proper spacetime coordinate system it is always the case that

$$\frac{\mathrm{d}x_4}{\mathrm{d}p} \neq 0. \tag{17}$$

Accordingly, along such a curve the proper time coordinate continuously increases or decreases. Since any timelike curve remains a timelike curve under a proper coordinate transformation, two world points along such a curve can never receive the same value of the time coordinate through a proper spacetime coordinate transformation. On the other hand, when x_4 is constant, that is,

$$\frac{\mathrm{d}x_4}{\mathrm{d}p} = 0, \tag{18}$$

one is dealing with a spacelike curve, a *Strecke* or a spatial extension between two world points, for which the first three inequalities of (15) are positive.[57]

The significance of these coordinate restrictions for the principle of causality is then clearly spelled out:

> So we see that the concepts of cause and effect lying at the basis of the principle of causality also in the new physics never lead to inner contradiction, as soon as we always take the inequalities (15) in addition to our fundamental equations; that is, we restrict ourselves to the use of *proper* space-time coordinates (1917, p. 58).

As further discussed in Section 8, it is not *nature* but the structure of our cognitive experience (in Kantian terms, of our faculties of sensibility and understanding) that leads to the requirement that we use proper coordinate systems—it has to do *not* with the possible objects of physics, which belong to the conceptual realm only, but with the possible objects of experience (physical facts as determined by measure threads and light clocks) as these are represented standing in causal relations within spatio-temporal empirical intuition.

The discussion of preserving causal order of events through use of proper coordinate systems concludes with the introduction of the so-called 'Gaussian coordinate systems'. Hilbert employs such a system later at the end of his paper when integrating the Einstein equations in the Schwarzschild situation of the gravitational field of a mass point at rest. Gaussian systems are local coordinate systems for the region in M of any point p lying on a hypersurface $S^{n-1} \subset M$ ($n = dim\, M$) according to which any point q in the forward evolution sufficiently close to the given point p lies on *a unique geodesic* (of length $< \varepsilon$) leaving p orthogonally from S^{n-1}. In such coordinate systems also it can never be the case that the points p and q can be transformed to the same hypersurface of simultaneity, and so Gaussian systems are always *proper* coordinate systems.[58]

6.6. The problem of univocal determination

In turning now to the central problem of univocal determination in the new physics, Hilbert returned to the principal theme linking the two communications: the consequences for causal determination issuing from Theorem I of his First Communication. As we have seen, Hilbert regarded this theorem as the 'guiding motive (*Leitmotiv*) for the construction of my theory'. Theorem I stated that, due to the general invariance of the world function stipulated in Hilbert's Axiom I, there are 4 identities obtaining between the Euler–Lagrange equations of motion derived from the action associated with the world function. As a result, any theory with a generally invariant action will have 4 fewer independent equations of motion than dynamical variables; if the number of these variables is *n,* the number of independent equations is $n-4$. The implication, clearly stated in the Proofs, but excised entirely in the

[57] Hilbert's inequalities (15) apply only to curves that are spacelike or timelike. But points on a *null* curve may be transformed to simultaneity by a coordinate transformation that is 'proper' in the sense that it preserves causal order on all *timelike* curves. In this case the 'reality relations' will be violated even though the transformation satisfies Hilbert's causality principle, and so the inequalities (15) are merely sufficient, but not necessary, to preserve causal ordering. See Renn & Stachel (1999, p. 80). Our thanks to a referee for reminding us of this point.

[58] Gaussian coordinates preserve the number of independent equations but manifestly reduce the number of 'physical' potentials, thus making the overdetermination obvious. Our thanks to a referee for this way of phrasing the utility of Gaussian coordinates for Hilbert's purposes.

published version of Hilbert's First Communication (see Section 4.2 above), is the *causal underdetermination* of the dynamical evolution of the fields represented by the generally invariant action. We reiterate here that the scope of Theorem I extends to any generally invariant theory, and is thus broader than either general relativity or Hilbert's theory itself. Indeed, the tension it revealed between general invariance and causal determination is a *mathematical result attained via the axiomatic method,* not a result of heuristic argumentation (as in the 'hole argument', of which more below in Section 7). Appropriate to the meta-logical concerns of that method, Theorem I exposed an apparent *inconsistency* between fundamental principles, each supposed to have unrestricted validity. Resolving this problem is therefore the common agenda for Hilbert's two papers on Foundations of Physics.

While the conflict between general invariance and causal determination is only implied, via Theorem I, in the published version of Hilbert's first paper (and accordingly downplayed in the literature), in his 1917 paper Hilbert nonetheless claimed (see the quotation below) that there he had 'especially stressed' this fact. One might conjecture that Hilbert had merely forgotten that any explicit reference to the failure of Cauchy determination for his fundamental equations had been excised from the Proofs, along with his non-covariant treatment of energy. However, it is more plausible to think otherwise. Hilbert's first resolution had been cast in terms of finding 'four additional non-invariant equations', a strategy that hadn't worked. Then, when revising the Proofs in the light of Einstein's work, it seems he had not yet seen that the solution lay not in four additional non-invariant equations, but rather in the four *inequalities* (15). Uncertain about how the issue was to be resolved, Hilbert had simply buried the entire issue in the published version.

In the event, the main point of the 1917 paper is to provide a quite different manner of resolution. Although continuing his interpretation of Theorem I that the four generalized Maxwell equations (3) are a consequence of the ten gravitational equations (2), this claim lies well in the background, while the matter of causality is given pride of place. The basic achievement of the paper is to give the necessary *reformulation* of the causality principle that is required by the new physics of general invariance.

The need for such a reformulation is explicitly stated. Hilbert observed that up until the present time all physical theories permitting a variational formulation have satisfied the requirement of causality, in the sense that this formulation yielded a determinate system of differential equations, providing univocal determination of future states from present states and their time derivatives. As precisely formulated by Cauchy, causal determination requires that the theory provide an independent equation for each unknown function appearing in the theory, a result secured by 'the well-known Cauchy theorem on the existence of integrals of partial differential equations'. However, the situation is otherwise once the requirement of general invariance is raised:

> Now the fundamental eqs. (2) and (3) set up in my first contribution are, as I especially stressed there, in no way of the above-characterized kind. Rather, according to Theorem I four are a consequence of the remaining ones: We viewed the four Maxwell equations (3) as a consequence of the ten gravitational equations (2) and therefore have only the 10 essentially independent equations (2) for the 14 potentials $g_{\mu\nu}$ and q_s.
>
> As soon as we raise the requirement of general invariance for the fundamental equations of physics, the just mentioned circumstance is even essential and necessary (1917, pp. 59–60).

On the other hand, Hilbert claimed that the situation in the newly emerged generally invariant physics is such that

> from knowledge of physical magnitudes in the present and past, it is no longer possible to univocally deduce their values in the future.

This contention is illustrated by the vivid (*anschaulich*) example of an imagined integration of his fundamental equations (2) and (3), yielding a solution corresponding to a single electron 'at rest'. (Of course, neither Hilbert nor anyone else actually obtained such solutions!) In this case, the 14 potentials are determinate functions only of the space coordinates, are completely independent of the time coordinate x_4, and as well, the first three components of the charge and current density (*Viererdichte*) may vanish. Hilbert then transforms the 14 potentials according to a coordinate transformation in which all quantities retain the old values in the primed system, except that for $x'_4 > 0$, one spatial coordinate (say, in the x direction) transforms as a function of x'_4. This is a coordinate transformation in the space of solutions that apparently transforms a resting electron into one in motion. But in allowing such a transformation to an *improper* coordinate system, two conflicting descriptions of the behavior of the electron are given—one in which it remains at rest, and the other in which it suddenly (*deus ex machina*) starts to move. Thus, the underdetermination at the level of the fundamental equations leads to a problem at the level of spatio-temporal—and causal—description.

As a result, Hilbert argued, we are driven to *reformulate* the causality principle through 'a sharper grasp' of how the general invariance of the new physics should be understood. The general invariance of the laws is, as we shall see further in Section 8, a regulative ideal of physical objectivity that applies to the conceptual structure of fundamental (field) physics. There remains the question of how the principle of general invariance should be understood not in the context of laws but in that of individual statements concerning the spatio-temporal evolution of particular systems or objects. Hilbert therefore revisited the question of what is meant by the meaningfulness of physical statements once the principle of causality is taken into account. His solution can be elucidated as follows. A *necessary* condition for such a statement to be physically meaningful is that it has a generally covariant formulation. But of course, this is not *sufficient*. For when such statements are predictions, i.e., concern the future, Hilbert stipulated that their meanings are to be understood in such a way that the requirement of physical causality (viz., that causes precede their effects) is satisfied:

> As now regards the principle of causality, the physical quantities and their time-rates of change may be known at the present time in any given coordinate system; then a statement will have physical meaning only when it is invariant with respect to all those transformations for which precisely those coordinates used for the present time remain unchanged. I affirm that statements of this kind for the future are all univocally determined, that is, *the causality principle holds in this formulation: From the knowledge of the 14 physical potentials* $g_{\mu\nu}$, q_s *in the present, all statements concerning them for the future follow necessarily and univocally, in so far as they have physical meaning* (1917, p. 61).

Renn and Stachel (1999, p. 81) correctly observe that this is obviously *not* a claim that physically meaningful statements are independent of the choice of a coordinate system. On the other hand, neither is it evidence for what they go on to suggest, that Hilbert still attaches 'some residual physical meaning to the choice of coordinates'. Rather, as is apparent from Hilbert's formulation, the criterion of physical meaningfulness of statements requires satisfaction of the principle of causality in the usual sense that conditions in the present determine those in the future. Furthermore, any such physical statement must be independent of how it is designated by coordinates; i.e., it must be, in Hilbert's terms, an invariant statement.[59]

With this conception of the causality principle in hand, we can formulate the necessary and sufficient conditions for a proposition to be physically meaningful:

(a) The proposition must have a generally covariant formulation.
(b) When the proposition is expressed with respect to a proper coordinate system, the truth value of that description must be *uniquely* determined from an appropriate spacelike past hypersurface.

In other words, when we express the propositions of physics in terms of possible objects of experience (that is, including the spatio-temporal and causal aspects of how we experience objects), those statements are physically meaningful if and only if they are causally determinate in the sense of condition (b), as well as satisfying condition (a).

From Hilbert's point of view, the physical principle of causality, as preserved by the coordinate conditions of a well-posed Cauchy problem, is a lingering but ineliminable constraint on human understanding ('physical meaningfulness'), a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense qualities, the requirement of physical causality is anthropomorphic, having to do not with the objective world of physics but rather with our experience of that world.[60]

Hilbert offers an existence proof for his reformulated principle of causality in terms of a Gaussian spacetime coordinate system; in such a system, corresponding to the 10 potentials of Hilbert's theory,

$$g_{\mu\nu}(\mu,\nu = 1,2,3);\quad q_s(s = 1,2,3,4) \tag{19}$$

is a system of just as many partial differential equations. Integrating them on the basis of the given initial values for $x_4 = 0$, then the values of (19) for $x_4 > 0$ are univocally determined. Expressed in these terms, any statement about the univocal determination of the future is of invariant character.

There are, furthermore, several diverse ways of mathematically expressing invariant statements. Three examples are cited. The first pertains to the choice of a particular

[59]These points are made explicitly in Hilbert (1916b), probably dated 21 November 1916 (see Sauer, 2001, p. 5), on pages 5–6: 'We will prove that the thus formulated causality principle: "All meaningful assertions are a necessary consequence of what has gone on before (*der vorangegangenen*)" is valid'. Also, page 5: 'That is, one must not only say that the world laws are independent of reference system, but rather also that any individual assertion regarding an occurrence or a coincidence (*Zusammentreffen*) of occurrences only has a meaning if it is independent of designation, i.e., if it is invariant'.

[60]Norton (1993, pp. 805–806) briefly discusses the electron example (see above), and claims that, according to Hilbert, statements about the motion of the electron are physically meaningless because they are not invariant, but one can make them invariant by the introduction of a Gaussian coordinate system. According to our exposition of Hilbert, physical meaningfulness depends on the satisfaction of two conditions: condition (a) is satisfied by some generally covariant statement about the electron's worldline, while condition (b) is satisfied via the restriction to Gaussian coordinate systems such that, when we start talking about the motion of the electron, we don't make statements that contradict one another concerning the causal history of the electron.

coordinate system in accord with the geometry of a physical situation, e.g., the use of polar coordinates in treating the spherically symmetric gravitational field of a mass point (Schwarzschild solution). The second way is by *elimination* of coordinates from any relations, so that a statement about the electron at rest in a *suitable* coordinate system is a physically meaningful statement. Finally, and most familiarly, a statement is physically meaningful if it is invariant under arbitrary transformations of the coordinates. Interestingly, Hilbert here cites the example of energy in general where the ('pseudo-tensor density') expression for the energy-momentum-stress of the gravitational field is not generally invariant but nonetheless, if defined properly, occurs in the statement of a conservation law that holds in every frame, i.e., is generally covariant.

In our opinion, Hilbert's main aim, and the central achievement, of the published version of his Second Communication on the Foundation of Physics is to provide this reformulation of the causality principle, appropriate for generally covariant physics.

6.7. *Physics as a four-dimensional pseudogeometry*

The final part of Hilbert's paper begins with the declaration that, according to his explanations,

> physics is a four-dimensional pseudogeometry whose metric determination $g_{\mu v}$ is bound, according to the fundamental equations (2) and (3) of my first contribution, to the electromagnetic quantities, that is, to matter (1917, p. 63).

This is a sweeping statement, and certainly a precursor of the even bolder declaration the next year of Weyl's 'theory of gravitation and electromagnetism':

> Everything real (*wirklich*) that transpires in the world is a manifestation of the world-metric. Physical concepts are none other than those of geometry (Weyl, 1918b, p. 385).[61]

Of course, in part, Hilbert's was a claim about the validity of Mie's electromagnetic theory of matter. But that is probably not the only fault that Einstein in 1916 would have found with it. For it is also the expression of a commitment to the kind of physics that the future would bring—a purely geometrical physics of fields under the requirement of general invariance. Einstein was still some years away from explicitly endorsing such a program (*sans* the Mie theory) in print. Beginning with his first papers on unified field theory in the early 1920s he was, by then, *de facto* committed to it, and it retained his allegiance to the end.

In any case, Hilbert's declaration serves the rather specific purpose of enabling him to raise the 'old geometrical question' of the validity of Euclidean geometry for reality, a question that the new physics of the fundamental equations (2) and (3) have rendered 'ripe for solution'. The setting is this. Previous physics presupposed a prior geometry for the formulation of its laws, whereas this is not done in the new physics ('of Einstein's principle of general relativity'). Proudly emphasizing his First Communication, Hilbert states that now the physical and geometrical laws, derived 'at one blow' from a variational ('Hamiltonian') principle, show how the geometrical quantities $g_{\mu v}$—also the mathematical expression for the phenomena of gravitation—are connected to the electrodynamic

[61] For discussion, see Ryckman (2005, chapters 4–6).

potentials q_s, and thus to the physics of matter. Hence in the new pure field physics there can be no question of a background geometry; physics and geometry have become *one* science resting on a common foundation.

In this setting of pure field physics, Euclidean geometry appears as 'an unnatural law of action-at-a-distance' (*ein fremdartiges Ferngesetz*), inconsistent with the next-to-next causal mechanism presupposed by all of field physics. Yet since the 'pseudo-Euclidean geometry' (i.e., Minkowski metric),

$$g_{11} = 1, \quad g_{22} = 1, \quad g_{33} = 1, \quad g_{44} = -1; \quad g_{\mu\nu} = 0 \to (\mu \neq \nu), \tag{20}$$

nevertheless appears in certain circumstances as a solution to the fundamental equations of the new physics, an investigation is needed as to whether, and under what conditions, these are the only regular (i.e., non-singular) solutions. The involved calculations of the last twelve and a half pages of Hilbert's paper is largely concerned with this task, and two specific problems are addressed.

6.7.1. An electricity-free world

Hilbert is first concerned to scrutinize his fundamental equations when the $g_{\mu\nu}$ adopt the 'pseudo-Euclidean' (Minkowski) values (20), and similarly the electrodynamic potentials $q_s = 0$. Setting the former into the first term of the fundamental gravitational equations (4)—see Section 4.1 above—we obtain

$$[\sqrt{g}K]_{\mu\nu} = 0, \tag{21}$$

while setting the latter into the second term of the fundamental equations (4) yields

$$\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{22}$$

This corresponds to the case of a world distant from all electromagnetic fields (and so from matter, in Hilbert's conception). Under such circumstances, pseudo-Euclidean geometry is possible, but is it necessary? To answer this question, Hilbert considers arbitrarily small variations of the Minkowski metric, according to a linear approximative procedure recently employed in Einstein (1916b) for the integration of the gravitational field equations.[62] Under certain natural assumptions about the perturbative terms, Hilbert arrived at the conclusion that indeed such solutions are the only 'regular' solutions:

> *Through variation of the metric of the pseudo-Euclidean geometry under the assumptions …, it is not possible to obtain a regular metric that is not patently pseudo-Euclidean and which yet at the same time still corresponds to an electricity-free world* (1917, p. 66; original emphasis).

In other words, under these assumptions, the Minkowski metric is the only non-singular metric corresponding to an electricity-free world. We recall that, on account of the principle of general invariance, in the early years of general relativity it was not immediately clear how an intrinsic characterization of a spacetime (curvature) singularity

[62] Though the paper contains mathematical errors later corrected, this is the first paper in which Einstein argued that (in the weak-field approximation) there exist gravitational waves, propagating at light velocity. See Pais (1982, pp. 279–280).

might be given.[63] In point of fact, Hilbert (1917, p. 70) opined that only regular solutions to the fundamental physical equations were immediately represented in reality. As we know, this was also one of Einstein's most pious beliefs (see Earman & Eisenstaedt, 1999). Nonetheless, Hilbert was the first to attempt to provide a definition of a non-singular or 'regular' metric. As implicitly assumed here, but explicitly stated in his winter semester lectures in 1916/1917, Hilbert regards a metric as 'regular' if there exists a system of coordinates in which the $g_{\mu\nu}$ are continuous and differentiable, with a non-zero determinant of the coordinate transformation, specified to be one–one and invertible.[64] However, subsequent work showed that Hilbert's definition of a non-singular spacetime metric is inadequate, not succeeding in distinguishing true curvature singularities from those arising from coordinate transformation (e.g., Earman, 1995, p. 6). This must be taken into account in his discussion of the Schwarzschild singularity.

6.7.2. *Rederivation of the Schwarzschild solution for an isolated resting mass point*

Prior to this in 1916, Schwarzschild had integrated Einstein's field equations for the gravitational field of a resting point-mass, corresponding to the sun's gravitational field, a central case for the empirical tests of general relativity. This was the first exact solution of the field equations. Still earlier in November 1915, and later, in 1916, Einstein had given an approximate solution for the same field. Both had made various assumptions about the quantities $g_{\mu\nu}$. Schwarzschild assumed that they had 'pseudo-Euclidean' values at infinity, i.e., that a 'pseudo-Euclidean' metric obtains everywhere outside the region of integration. The gravitational field of a point-mass at rest in Einstein's 'approximative integration' was, on the other hand, calculated as a small perturbation of the metric of Minkowski (i.e., 'pseudo-Euclidean') geometry. Hilbert concluded his paper by re-deriving the Schwarzschild exact solution without any assumptions about the $g_{\mu\nu}$ at infinity, hence without the implicit assumption of the global validity of pseudo-Euclidean geometry. As we know, Einstein (1917) achieved the same result in explicit cosmological terms with the concept of a finite but unbounded universe.

6.8. *Conclusion concerning the Second Communication*

Our contention is that the material cut from the Proofs establishes a thematic linkage between the First and Second Communications. Specifically, the material cut from the Proofs concerns Hilbert's first attempt to resolve the tension between causality and general covariance, precisely formulated by Theorem I. In the 1917 published version of the Second Communication, Hilbert offers us a detailed discussion of this tension, including a new resolution of the 'problem of causality', and a discussion of to what extent pseudo-Euclidean ('action-at-a-distance') geometry is present in the new theory. The setting for the Second Communication is Hilbert's axiomatic method, as it was in the First Communication. In our opinion, Hilbert's Second Communication is an integral and natural continuation of the project set up in his First Communication, and contains important and interesting developments of that project. It is not, as has been affirmed, a

[63] Earman (1995) emphasizes that such a characterization (as 'geodesic incompleteness') came only in the 1960s in the work of Penrose and Hawking.

[64] Winter Semester 1916/1917 lectures, *'Die Grundlagen der Physik II'*, p. 118. See the discussion in Eisenstaedt (1989, p. 218), and in Renn & Stachel (1999, pp. 87–88).

hodge-podge of unrelated issues presented within the framework of Einstein's completed theory. Indeed, recapitulating the essential details of the theory presented in his First Communication in lectures entitled 'The World Equations' in Hamburg in the summer of 1923, Hilbert, so far from being in the 'agony of facing the collapse of his own research program', considered Einstein's concurrent 1923 publications on unified field theory as having arrived, after a 'colossal detour' ('*kolossaler Umweg*'), at a '*Hamiltonsche Prinzip*' (Lagrangian density for the combined gravitational and matter fields) essentially similar to that Hilbert originally proposed in December 1915.[65] It is, presumably, this convergence that led Hilbert to republish, in 1924, both of his communications in revised form in the *Mathematische Annalen.*

7. Hilbert and the 'hole argument'

The tension between general covariance and causality is just the problem encountered by Einstein in the 'hole argument'. But in Hilbert's hands, is there any reason to think that, as with Einstein, the difficulty arose from wrongly attributing physical meaning to the spacetime coordinates? That Hilbert's four non-covariant energy equations do implicate such a misunderstanding is affirmed by Renn and Stachel (1999, pp. 77, 83):

> … we will discuss Hilbert's treatment of the problem of causality in Paper 2 [Hilbert (1917)] and encounter further evidence for his neglect of Einstein's insight that, in general relativity, coordinate systems serve as mathematical devices for the description of space-time coincidences and have no physical significance of their own. … In summary, Hilbert's treatment in Paper 2 of the problem of causality in general relativity still suffers from the flaws of his original approach, in particular, the physical significance he ascribed to coordinate systems and his claim that the identities following from Theorem I represent a coupling between two sets of field equations.

We could not disagree more strongly with everything said in these remarks. Certainly, there is a construction in Hilbert's so-called 'Causality Lecture' (1916b) that looks superficially similar to the 'hole argument', but the relevant question is what problem Hilbert saw as following from his construction. It seems clear to us that the issue concerning Hilbert was very different from that which plagued Einstein, and we have three sources of supporting evidence.

Consider first the direct evidence: Hilbert's Theorem I as compared with Einstein's 'hole argument'. These can be considered as different construals of the same problem (causal underdetermination) *just to the extent that* if (counterfactually) the Euler–Lagrange equations in a generally covariant theory were all independent, then a 'hole argument' construction could be given which would show a failure of univocal determination.[66] But of course, this is just what Hilbert's Theorem I prohibits. There is therefore no reason to think Hilbert accepted the conclusion that hindered Einstein for two long years, that a

[65]See Hilbert (1923), and Majer & Sauer (2004, p. 263), who state that Hilbert identified the differences with Einstein 'as being of a purely nominal nature'.

[66]As Brown & Brading (2002), p. 70, state: 'these considerations represent the flipside of the *underdetermination* problem that had caused Einstein such headaches … the 'hole argument'…. In a sense they are one and the same problem'.

generally covariant gravitational theory was not possible because of failure of causal determination. After all, even with his four supplementary energy equations, the field equations of Hilbert's gravitational theory, even if only written in abbreviated form as Lagrangian derivatives, are generally covariant already in the Proofs, as seen above in Section 4.1. And although Hilbert may well have been aware of Einstein's difficulties in the 'hole argument' (see e.g., Howard & Norton, 1993), his analysis of the tension between general covariance and causality began with the fact that the Euler–Lagrange equations arising from his variational problem are not all independent—*as must be supposed to become entangled in the snares of the* '*hole argument*'. Rather, Theorem I, stating that four linear combinations of the Euler–Lagrange brackets are divergences (now usually called the generalized Bianchi identities), posed an immediate difficulty for achieving Cauchy determination in generally covariant theories. Hilbert's discussion of causality within the frame of the Cauchy problem (and to a lesser extent, the initial value problem) laid out an entirely different route than that adopted by Einstein in the 'hole argument' for resolving the problematic relation between general covariance and causality. There is no physical or metaphysical hypothesis involved (e.g., manifold substantivalism or determinism), nor any confusion about the physical significance of spacetime coordinates.

Second, there is Hilbert's strategy, exactly opposite to Einstein's, for resolving the apparent tension between general covariance and causality. Unlike Einstein, Hilbert never contemplated abandoning general covariance. It is an overriding fact that an *axiom* of general invariance guided Hilbert's construction, and that surrendering general invariance in the face of an apparently irremediable conflict with causality was, apparently, never an option. Although Hilbert was not yet sure how, or even if, this tension might be resolved, he gave many indications that if need be the Gordian knot was to be cut by surrendering causality—in the sense of Cauchy determination—not general covariance.[67] The four non-covariant additional conditions that Hilbert initially sought therefore did not so much break the covariance of his theory as show what additional structure must be added to a generally invariant theory in order to render it compatible with univocal determination. On the other hand, the strategy of Hilbert (1917) expressly deals with the causal principle in the sense that Kant maintains is intrinsic to the nature of human cognition. As we have shown, this is in accordance with Hilbert's general philosophical view that physical causality, and so irreversibility, is a phenomenon rooted exclusively in the human perspective on the world (e.g., Hilbert, 1919–1920, pp. 86–87). Once the axiomatic method had mathematically demonstrated a *prima facie* antagonism between two pillars of objectivity, one going back to Kant (physical causality) and one reaching beyond Kant (general invariance), Hilbert sought a harmonious remedy that would expand the sphere of *what is a possible object of physics* in accord with general invariance yet must be compatible with our experience—i.e., we obviously must be able to recover an account of possible objects of experience, of physical facts read off from measure threads and light clocks, on pain of failure of empirical meaning, or worse, empirical incoherence. In short, Hilbert never considered giving up general covariance as an axiom (clearly this is not a possible solution to the problem he saw himself as facing), but rather sought his solution in finding what further conditions must

[67] See the 'Causality Lecture' (1916b), and the Winter Semester 1916/1917 lectures, '*Die Grundlagen der Physik II*'.

supplement a generally covariant theory enabling us to recover possible objects of experience.

In our opinion there is no reason to think that the problem Hilbert saw himself as facing with respect to causality arose, as it did in Einstein's case, from wrongly attributing physical meaning to the spacetime coordinates. Hilbert's problem was a different problem, and one that the 'point coincidence' solution to Einstein's 'hole argument' would not even touch upon.

Finally, there is the contextual evidence highlighted by Howard and Norton (1993), who revealed that there was a 'Göttingen Answer to the hole argument', probably based on (as Felix Klein put it, in July 1915) the widespread familiarity of Göttingen mathematicians with the use of arbitrary coordinates in the work of Lagrange, Gauss, and Riemann. Without paying undue regard to Hilbert's stature as perhaps the world's premier mathematician in 1915, it is almost inconceivable that Hilbert suffered from confusions on an issue clearly grasped by the rest of the Göttingen mathematical community. As evidence, we simply underscore again Hilbert's remarks, in both Proofs and published papers, that coordinates are arbitrary designations of physical events, and his use of the term 'world parameters' to highlight this arbitrariness.

To think that Hilbert could have been a victim of Einstein's 'hole argument' is, we maintain, to misunderstand his project entirely.

8. Hilbert's revision of Kant in the light of general invariance

In the context of a discussion of Hilbert's views on foundations of mathematics, Peckhaus (1994, p. 91) has recently remarked that

> Hilbert's preference for Kantianism was largely incompatible with Kant's philosophy.

However, this statement should be qualified, for its truth largely depends on what is meant by 'Kant's philosophy'. To many of the neo-Kantians active in the first quarter of the twentieth century, Kant's philosophy was largely a work in progress, not a finished edifice, and nearly all were prepared to concede, as was Hilbert (1930, p. 383), that 'Kant greatly overestimated the role and the extent of the *a priori*'. By the same token, it might be said that Hilbert, through the axiomatic method, was the first neo-Kantian to put his finger on exactly where the general theory of relativity required a modification in the traditional Kantian transcendental framework that expressly bound considerations of objectivity together with conditions of possible experience. In Kant, space and time, as subjective forms of sensibility, are at once also objective conditions for perception of objects—conditions of the possibility of experience. For a cognition to be *objectively valid* (to be a representation pertaining to a possible object *for us,* hence to be *meaningful*) is for it to invoke *our* specifically human type of finite, receptive spatiotemporal sensory intuition of objects.

Hilbert essentially argued that these Kantian conditions are no longer fully universal once the requirement of general invariance is imposed on fundamental physical theory. While retaining part of Kant's tie of conditions of physical *meaningfulness* to *sensibility*, Hilbert posited general invariance as an axiom, a superordinate criterion of physical objectivity, attributing this development expressly to the influence of Einstein's gravitational theory. Lecturing in the winter of 1919/1920, he affirmed:

> Einstein proceeds from the demand that all laws of nature must be *invariant with respect to arbitrary transformations*; in passing from one coordinate system (x,y,z,t) to another through an arbitrary transformation
>
> $$\begin{aligned} x' &= X(x,y,z,t) \\ y' &= Y(x,y,z,t) \\ z' &= Z(x,y,z,t) \\ t' &= T(x,y,z,t) \end{aligned}$$
>
> (satisfying only certain conditions of continuity, differentiability, and invertibility), the mathematical form of the laws of nature should remain unaltered.
>
> The sense of this general 'principle of relativity' is that a representation of nature's linkages (*Naturzusammenhänge*) once and for all free of subjectivity and arbitrariness can only be such as is independent of the way in which the world points are designated (through coordinates) (1919–1920, p. 49).

For this the 'axiomatic method' is ideally suited, for it works up a domain of objects concretely given to sensibility into a purely logical-formal system of relations. Space and time (and causality, presupposing both) are necessary moments of human cognition, required also as conditions of the possibility of observation and measurement in science. Though they remain conditions of the possibility of experience, representation in space and time reflects merely the subjective origin of cognition in sensible experience; in the new physics, such representation is no longer a necessary criterion of *being a possible object of physics*. There, Hilbert made it clear, the pull of objectivity leads *away* from intuition, from physical observation and measurement, and indeed from everything *anthropomorphic*. This, of course, is just the direction of cognition captured within the axiomatic method.

The new physics of general invariance retains the traditional goal of physical science: to achieve a completely objective description, a systematically unified, observer-free cognition of nature. But it also has shown that this requires a further emancipation (never completely attainable) of our conceptions of nature from all the subjective elements pertaining to human sensibility. The principal step along this new path is just Einstein's requirement of general covariance.

> Hitherto, the objectification of our view of the processes of nature took place by emancipation from the subjectivity of human sensations. But a more far reaching objectification is necessary, to be obtained by emancipating ourselves from the *subjective* moments of human *intuition* with respect to space and time. This emancipation, which is at the same time the high-point of scientific objectification, is achieved in Einstein's theory, it means a radical elimination of *anthropomorphic* slag (*Schlacke*), and leads us to that kind of description of nature which is *independent* of our senses and intuition and is directed purely to the goals of objectivity and systematic unity.[68]

In referring to 'Einstein's theory', it is clear that Hilbert regarded general covariance as its philosophical focal point, as a supreme 'principle of objectivity'. A generally covariant

[68] Hilbert, 1921, *Grundgedanken der Relativitätstheorie*, Summer Semester 1921 lectures, edited by P. Bernays; cited by Majer, 1995, p. 284.

representation of nature, as Hilbert stated in lectures in Hamburg in 1923, is one of the triumphs of the human mind:

> The (general) principle of relativity signifies, it seems to me, for the first time a definitive, exact and general expression concerning the laws obtaining in reality (*Wirklichkeit*), and accordingly stands, in my opinion, *at the pinnacle* of pure achievements of thought of the human mind (*steht somit meiner Meinung nach unter den reinen Gedankentaten des menschlichen Geistes* ***oben an***) (Hilbert, 1923, lecture 1, p. 16).

In accordance with 'the general principle of relativity' (i.e., general covariance), physical science should aspire to a *non-anthropomorphic* account of physical meaningfulness and objectivity, to a 'description of nature which is *independent* of our senses and intuition'.

Yet, accorded axiomatic status by Hilbert, the principle of general invariance *is neither true nor false, but a regulative idea*,

> if, in accordance with Kant's words, we understand by an idea a concept of reason that transcends all experience and through which the concrete is completed so as to form a totality (Hilbert, 1926, p. 392).

According to Kant, such ideas are the product of the faculty of reason, not of the understanding. Here we may recall that the epigram Hilbert chose for *Grundlagen der Geometrie* occurs at the conclusion of the 'Transcendental Dialectic'. It is the pithy summary of what the 'Transcendental Dialectic' seeks to show, as stated at the very beginning, that the character of human cognition is to be understood as interrelating the distinct contributions of three faculties of mind.

> All our cognition starts from the senses, goes from there to the understanding, and ends with reason, beyond which there is nothing higher to be found in us to work on the matter of intuition and bring it under the highest unity of thinking (A298-299/B355).

Accordingly, the central goal of the 'Transcendental Dialectic' is to complete the account of cognition presented in the 'Transcendental Analytic' (the first part of the *Critique of Pure Reason*, pertaining to sensibility and the categories of the understanding) by adding precisely the cognitive role of 'theoretical reason' in its 'critical' employment. In general, Kant viewed reason (both 'theoretical' and 'practical') as rooted in the human capacity to project beyond given experience to seek the totality of possible experience, the 'totality of all conditions' or indeed the 'unconditioned' presupposed by any series of conditions. Of course, such a totality can never be an object of possible experience, and so cannot be considered an object of cognition; this is the conclusion of the arguments of the Antinomies of Pure Reason. But the regulative ideas of theoretical reason express what is nonetheless essential to natural science: reason's capacity to surpass the confines of experience through the hypothetical adoption of maxims of systematic unity or 'unity of nature'.[69] According to Kant, this

> systematic unity (as mere idea) is, however, only a *projected* unity, which one must regard not as given in itself, but only as a problem (A647/B675).

[69]See Buchdahl (1969) and Neiman (1994) for forceful arguments that regulative ideas are essential to Kant's conception of science, in particular that reason is essentially involved in any cognition of 'the order of nature'.

The ideal concepts or principles of such unity are then the product of the 'hypothetical use of reason, on the basis of ideas as problematic concepts', and are 'not properly *constitutive*' (for 'constitutive' in Kant's sense pertains only to objects of experience). As Majer (1993a) in particular has emphasized, by employing a 'method of ideal elements' in his proof-theory (regarded as a strictly finite means to prove the consistency of transfinite mathematics), Hilbert expressly tied the use of ideal elements in mathematics to Kant's regulative use of ideas. While Hilbert allowed that such elements could be introduced either through mathematical construction or as axioms,[70] the axiomatic method, naturally, prefers the latter. In virtue of their ideality, and so severance from experience and intuition, axioms play a hypothetical role in cognition, no longer constitutive of experience but informing and guiding the idea of physical objectivity.

Thus for Hilbert, the principle of general covariance has shown that the objective scientific description of nature must transcend experience, even as it begins with experience and so with sensibility. Experience requires that events are related as cause and effect: for Hilbert, the principle of causality remains synthetic *a priori*—a condition of possible experience in Kant's sense. However, as we have seen, there is a tension between causality and general covariance. Theorem I follows (*inter alia*) from an axiom of general covariance, which Hilbert viewed as a necessary requirement of physical objectivity, a postulate governing any fundamental physical theory purporting to describe a mind-independent nature. But, according to Theorem I, in any such theory an apparent problem of causal underdetermination immediately arises (see Section 4.2). Hilbert emphasized that in physical theories prior to general relativity, including special relativity, the initial data can be freely chosen without explicit regard to special constraints and the relation of causality is fully determinate *because of the fixed background structure of space*, *and the unique global direction of time*. In generally covariant physics this is no longer possible once spacetime geometry becomes a dynamical component of the physical theory. Hilbert's resolution is to claim that the causal principle, and sensible representation in general, is a bit of 'anthropomorphic slag' that ideally can be jettisoned from the objective description, but not, on pain of incoherence, from what is given in physical experience as the determination of future states from past and present ones. Thus, as Hilbert would come to understand the matter, in any generally covariant theory the initial data must be specified appropriately so as to allow the univocal determination of future states from present states. Specifically, the constraint of causality, expressed as the four inequalities (15) for the $g_{\mu\nu}$, required that events on a given timelike worldline always stand in an unambiguous causal relation. So these inequalities for the $g_{\mu\nu}$, essentially specifying that x_4 is timelike, in fact provide a normal system of equations for the remaining gravitational potentials.

On this approach, sensible representation (and so representation of causal relations in space and time) remains as a condition of possible physical experience (ultimately reducible to measurements with 'measure threads' and 'light clocks'). It is not nature, but the nature

[70]E.g., Hilbert (1919–1920, p. 91): 'The transition to the broader system [containing ideal elements] either can be effected in a constructive manner, in which the new elements are developed through mathematical construction from the older ones, or in an axiomatic manner, in which the new system is characterized through relational properties. In the second case, it requires a proof that the supposition of a system with the desired condition does not in itself lead to contradiction'.

of our sensibility, that mandates the requirement of 'proper coordinate systems' in restoring univocal determination of the future from the present. Such coordinate systems nonetheless enable objective assertions about the processes of nature, so long as there exists a generally covariant formulation of these assertions (see Section 6.6, above). But the requirement of causality is no longer a condition governing the ideal conception in physics of a mind-independent world as the object of physical inquiry.

This completes Hilbert's revision of the Kantian *a priori*. In the 1920s, Hilbert termed his revised understanding of the *a priori* 'the finite viewpoint', taking from Kant the methodology or standpoint that objective cognition can only be understood as conditioned by *a priori* structures of the mind, but refashioning the boundaries of the *a priori* somewhat differently:

> The *a priori* is thereby nothing more and nothing less than a basic viewpoint or expression for certain essential preconditions of thought and experience. However, the boundaries between that we possess *a priori* on the one hand, and that, on the other, for which experience is necessary, we must draw differently than did Kant; Kant far overestimated the role and the scope of the *a priori.*
>
> We see therefore: in the Kantian theory of the *a priori* (*Apriori-Theorie*) there is still contained anthropomorphic slag (*Schlacken*), from which it must be freed, and after such removal only that *a priori* viewpoint (*apriorische Einstellung*) is left, which also lies at the foundation of pure mathematical knowledge: it is essentially that finite viewpoint characterized by me in different essays (Hilbert, 1930, pp. 383, 385).

In perhaps a hitherto unsuspected manner, Hilbert's 1915–1917 treatment of gravitation and electromagnetism within the frame of the axiomatic method provides new evidence for Majer's (1993b) claim that

> Hilbert's 'finite point of view' is not restricted to mathematics or meta-mathematics, but is stated as a universal principle of epistemology (p. 191).

9. Conclusions

Theorem I of Hilbert's First Communication, based on the *axiom of general invariance*, provided a purely mathematical diagnostic of the tension between causality and general covariance appearing as a problem of causal underdetermination. The material cut from the Proofs is his first attempt at resolving this tension.

The Second Communication (1917) presents Hilbert's later, and different, attempt at a resolution. For Hilbert, general covariance is a refinement, first identified as such by Einstein, of the condition of objectivity that must be placed on fundamental physical theory, a necessary casting off of the 'anthropomorphism' of particular observers and particular points of view. Its status is that of an ideal regulative principle in the sense of Kant, governing the image constructed in physical theory of an objective world that is no longer adequately representable through the senses or by intuition. Hilbert exalts this new requirement on objectivity by according it axiomatic standing.

Causality, in the physical sense of causal ordering, as other texts of Hilbert make clear, is *no longer* itself a principle of objectivity, although it remains a presupposition of measurement and so, a 'condition of possible experience'. That it is merely a subjective condition is shown by Hilbert's posit of the time-reversal invariance of the fundamental physical laws. The principle of causality, as preserved by the coordinate conditions of a well-posed Cauchy problem, is a lingering but ineliminable constraint on human understanding, a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense qualities, this requirement of causality is anthropomorphic, having to do not with the objective world but only with our experience of that world. While the aim of physical science is to eliminate such 'anthropomorphic slag' from its objective descriptions of the world, this goal may not be attainable because of the limits of human cognition.

The strategy of Hilbert's Second Communication is to adopt a criterion of physical meaningfulness and objective validity in terms of general covariance, while recognizing that human experience (measurement and observation) requires a restricted representation of space and time for which the causal relation is preserved. This reconciliation, set in the frame of the Cauchy problem, produced the four coordinate conditions employed to restore univocal determination of future states from present ones, and the requirement of causality is accordingly modified.

Hilbert's approach to, and resolution of, the tension between general covariance and causality was significantly different from that of Einstein, and it essentially resulted in a novel amendment of Kantian epistemology of science. Whereas Einstein, distraught at an apparent abandonment of causality (unique dynamical evolution of the gravitational field from given sources) required by the 'hole argument', abandoned general covariance for over two years, Hilbert, in the face of the apparent causal underdetermination (diffeomorphic freedom) of the new generally covariant field physics, considered taking just the opposite course of action. Having given general covariance axiomatic status, and then, on the basis of his Theorem I, observing that the Cauchy problem was not well posed, Hilbert contemplated surrendering causality (in the sense of Cauchy determination). Then, when Hilbert found a way to restore causality in the face of general covariance (essentially by imposing gauge conditions), he indeed subordinated causality—as a condition of possible experience—to general covariance, a superordinate condition of physical objectivity.

In sum, Hilbert's two communications on 'Foundations of Physics' are a complex epistemological response to general relativity, hitherto unrecognized, and directly tied to his employment of the axiomatic method.

Acknowledgments

We would like to thank Carlos Casanova, Sol Feferman, Brandon Fogel, Michael Friedman, Ulrich Majer, Brian Pitts, Tilman Sauer, and Joseph Zepeda, for their comments and suggestions. We are especially grateful to two anonymous referees for their detailed comments and advice. Material based on this research was presented at the Philosophy of Physics Research Seminar, Oxford University (2002); HOPOS (2004); the Logical Methods Workshop, Stanford University (2005); the Belgian Society for Logic and Philosophy of Science, Brussels (2005); the History and Philosophy of Science Department, Bloomington (2006); and at the 2007 conferences 'Philosophical and Formal Foundations of Modern Physics', Les Treilles, France, and 'Kant and the Philosophy of Science Today', University College, London; we are grateful to the organizers of these events for the opportunity to present this work, and to those present for their comments. In addition, one of us (K.B.) thanks Wolfson College, Oxford, for their support during the early years of the research behind this paper. We would also like to thank Penelope Brading for the application of her copy-editing skills as we prepared the final manuscript.

Appendix A. A sketch of the Cauchy problem in general relativity

This sketch is based on Adler, Bazin, and Schiffer (1975).

On Σ we have $x^4 = 0$. Taking the case of the gravitational equations for the vacuum ($R_{\mu\nu} = 0$) for sake of simplicity, the initial data on Σ are the functions

$$g_{mn},\ \ g_{mn},_4\ \ (m, n = 1, 2, 3) \tag{A.1}$$

of the three space variables x^1, x^2, x^3. Differentiating within Σ, the initial data also determine the quantities

$$g_{mn},_\mu \quad g_{mn},_{\mu\nu} \quad g_{mn},_{4\mu}. \tag{A.2}$$

Expressions (A.1) and (A.2) are therefore the Cauchy initial data. The vacuum equations $R_{\mu\nu} = 0$ can then be used to compute the second time derivatives of the metric potentials in terms of this known data, yielding the system of equations

$$R_{mn} = \frac{1}{2} g^{44} g_{mn},_{44} + M_{mn} = 0 \tag{A.3a}$$

$$R_{m4} = \frac{1}{2} g^{4n} g_{mn},_{44} + M_{m4} = 0 \tag{A.3b}$$

$$R_{44} = \frac{1}{2} g^{mn} g_{mn},_{44} + M_{44} = 0, \tag{A.3c}$$

where m, n = 1,2,3, and the $M_{\mu\nu}$ can be expressed in terms of the initial data on Σ (see further below). The spatial derivatives (indices m, n = 1,2,3) appear quite differently from the time index 4, and, as one can see, this system of second-order linear equations does not contain the unknown second time derivatives $g_{\mu 4},_{44}$ (μ = 1–4) needed to describe the time evolution of the metric from the data on Σ. Accordingly, the Cauchy problem in general relativity is *underdetermined* because the gravitational field equations do not give $g_{\mu 4},_{44}$ in terms of the initial data. Because of general covariance, this behavior is to be expected, since the coordinates off Σ can be transformed arbitrarily, and so unique functions $g_{\mu\nu}$ (x^1, x^2, x^3, x^4) cannot be obtained from the initial data. In particular, the $g_{\mu 4},_{44}$ can be chosen freely, the different choices corresponding to different coordinate systems but describing the same spacetime geometry. It is to remedy this underdetermination that Hilbert prescribes the use of 'proper coordinate systems', coordinates sufficiently rigid to eliminate the ambiguities due to coordinate freedom. In particular, we shall see that Hilbert adopted 'Gaussian coordinates' (now termed Gaussian normal coordinates) that, in placing the restrictions

$$g_{\mu 4} = 0,\ \ g_{44} = 1, \tag{A.4}$$

on the g_{mn} throughout spacetime, ensure that the second derivatives (A.4) are always zero. In effect, this means that a coordinate system is physically acceptable iff one of its variables x^4 is timelike and the other three spacelike, so that $g^{44} > 0$. (See further below.)

But the Cauchy problem is also *overdetermined*. The system of equations (A.3) consists of 10 equations for the six unknowns $g_{mn},_{44}$ on Σ itself, and so compatibility requirements must be imposed on the data $M_{\mu\nu}$ on Σ. Those restrictions are found by combining the equations (A.3a) with those of (A.3b) and (A.3c), and then, by reference to the vacuum equations in terms of the Einstein tensor $G_{\mu\upsilon} = R_{\mu\upsilon} - (1/2)Rg_{\mu\upsilon} = 0$, rewriting the system of equations (A.3) in the form

$$R_{mn} = 0, \tag{A.5a}$$

and

$$G^4_\mu = G^4_\mu = 0. \tag{A.5b}$$

Then the initial data cannot be given arbitrarily on Σ but must satisfy the four 'constraint equations' (A.5b). These four gravitational field equations are independent of the $g_{\mu\upsilon},_{44}$ and must be solved for the hypersurface Σ before attempting the Cauchy problem.

An attempt to construct a unique solution to that problem can now be plausibly summarized as taking place in three stages.

(i) On the initial spacelike surface Σ ($x^4 = x'^4 = \text{const.}$), the functions $g_{mn}(x^1,x^2,x^3)$ and $g_{mn'4}(x^1,x^2,x^3)$ are given, satisfying the four constraint equations $G^4_\mu = 0$. Four functions $g_{m4'44}(x^1,x^2,x^3)$ are chosen arbitrarily, while the six functions $g_{mn},_{44}(x^1,x^2,x^3)$ are determined by the vacuum equations $R_{\mu\upsilon} = 0$.

(ii) The quantities g_{mn}, $g_{mn'4}$ and $g_{mn},_{44}$ on Σ now determine $g'_{mn}\,(x^1, x^2, x^3)$, $g'_{mn},_4\,(x^1, x^2, x^3)$ on the neighboring hypersurface $\Sigma'(x^4 = x'^4 + \mathrm{d}x'^4)$ as

$$g'_{mn} = g_{mn} + g_{mn},_4\,\mathrm{d}x'^4$$

and

$$g'_{mn},_4 = g_{mn},_4 + g_{mn},_{44}\,\mathrm{d}x'^4.$$

These quantities are the initial data on Σ'. The constraint equations (A.5b) are automatically satisfied on Σ' since, if they are satisfied on Σ, then $G^n_m = 0$, and $G^n_m,_4 = 0$, so

$$G'^4_m = G^4_m + G^4_m,_4\,\mathrm{d}x'^4 = 0.$$

The four second derivatives $g'_{m4},_{44}$ can again be chosen arbitrarily on Σ', while the field equations $R'_{\mu\upsilon} = 0$ determine the $g'_{\mu\upsilon},_{44}$.

(iii) Iterate this process for each successive neighboring hypersurface.

Appendix B. Chronology

Summer 1915:	Einstein visits Göttingen and lectures on his approach to finding a new theory of gravitation
20 November 1915:	Hilbert presents his 'first communication'
25 November 1915:	Einstein presents the generally covariant field equations of his general theory of relativity to the Prussia Academy of Sciences in Berlin
2 December 1915:	Einstein field equations published
4 December 1915:	Hilbert's first presentation of his 'second communication'
6 December 1915:	Date of proofs of Hilbert's 'first communication'
26 February 1916:	Hilbert's second presentation of his 'second communication'

31 March 1916: Hilbert's 'first communication' published
29 January 1917: Hilbert's 'second communication' published

References

Adler, R., Bazin, M., & Schiffer, M. (1975). *Introduction to general relativity* (2nd ed.). New York: McGraw-Hill.

Born, M. (1914). Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie, Nachrichten von der Königliche Gesellschaft der Wissenschaften zu Göttingen. *Mathematisch-Physikalisch Klasse*, *12*, 23–36.

Born, M. (1922). Hilbert und die Physik. *Die Naturwissenschaften*, *10*, 88–93.

Brading, K. (2005). A note on general relativity, energy conservation, and Noether's theorems. In A. J. Kox, & J. Eisenstaedt (Eds.), *The universe of general relativity (Einstein Studies: Vol. 11)* (pp. 125–135). Boston/Basel/Berlin: Birkhäuser.

Brown, H., & Brading, K. (2002). General covariance from the perspective of Noether's theorems. *Diálogos*, *79*, 59–86.

Buchdahl, G. (1969). *Metaphysics and the philosophy of science: The classical origins from Descartes to Kant*. Oxford: Blackwell.

Corry, L. (1999). Hilbert and physics (1900–1915). In Gray (1999, pp. 145–188).

Corry, L. (2004). *David Hilbert and the axiomatization of physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Corry, L., Renn, J., & Stachel, J. (1997). Belated decision in the Hilbert–Einstein priority dispute. *Science*, *278*, 1270–1273.

Earman, J. (1995). *Bangs, crunches, whimpers, and shrieks: Singularities and acausalities in relativistic spacetimes*. New York: Oxford University Press.

Earman, J., & Eisenstaedt, J. (1999). Einstein and singularities. *Studies in History and Philosophy of Modern Physics*, *30B*, 185–235.

Earman, J., & Glymour, C. (1978). Einstein and Hilbert: Two months in the history of general relativity. *Archive for History of Exact Sciences*, *19*, 291–308.

Einstein, A. (1915). Die Feldgleichungen der Gravitation, *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte.* 844-7. Reprinted in Einstein (1996, pp. 245–248).

Einstein, A. (1916a). Die Grundlage der allgemeinen Relativitätstheorie, *Annalen der Physik* 49, 769–822. Reprinted in Einstein (1996, pp. 284–337). Translated by W. Parrett & G. Jeffrey as *The foundation of the general theory of relativity*. In Lorentz et al. (1923, pp. 109–165).

Einstein, A. (1916b). Näherungsweise Integration der Feldgleichungen der Gravitation, *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse,* 688–696. Reprinted in Einstein (1996, pp. 348–356).

Einstein, A. (1917). Kosmologische Betrachtunen zur allgemeinen Relativätstheorie, *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse,* 142–152. Reprinted in Einstein (1996, pp. 541–552). Translated by W. Parrett & G. Jeffrey as *Cosmological considerations on the general theory of relativity*. In H. Lorentz et al. (1923, pp. 175–188).

Einstein, A. (1921). Geometrie und Erfahrung, *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse*, pp. 123–130. Issued separately in expanded form, Berlin: J. Springer. Reprinted in Einstein (2002, pp. 382–405). Later translated by W. Parrett & G. Jeffrey as *Geometry and experience*. In *Sidelights on relativity.* New York: E.P. Dutton, 1923, pp. 235–256. Reprint New York: Dover, 1983, pp. 27–56.

Einstein, A. (1936). Physik und Realität. *Journal of the Franklin Institute*, 221, 313–337. Translated by S. Bargmann as *Physics and reality*. In Einstein, A. *Ideas and opinions,* NY: Crown Publishers, pp. 290–323.

Einstein, A. (1993). In M. Klein, A. Kox, & R. Schulmann (Eds.), *The Collected Papers of Albert Einstein: Vol. 5*. Princeton: Princeton University Press.

Einstein, A. (1995). In M. Klein, A. Kox, J. Renn, & R. Schulmann (Eds.), *The Collected Papers of Albert Einstein: Vol. 4*.

Einstein, A. (1996). In M. Klein, A. Kox, & R. Schulmann (Eds.), *The Collected Papers of Albert Einstein: Vol. 6*. Princeton: Princeton University Press.

Einstein, A. (1998). In R. Schulmann, A. Kox, M. Janssen, & J. Illy (Eds.), *The Collected Papers of Albert Einstein: Vol. 8A*. Princeton: Princeton University Press.

Einstein, A. (2002). In M. Janssen, R. Schulmann, J. Illy, C. Lehner, & D. Buchwald (Eds.), *The Collected Papers of Albert Einstein: Vol. 7*. Princeton: Princeton University Press.

Einstein, A., & Grossmann, M. (1913). *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig und Berlin: B.G. Teubner Reprinted in Einstein (1995), pp. 303–339.

Eisenstaedt, J. (1989). The early interpretation of the Schwarzschild solution. In D. Howard, & J. Stachel (Eds.), *Einstein and the history of general relativity (Einstein Studies: Vol. 1)* (pp. 213–233). Boston/Basel/Berlin: Birkhäuser.

Geroch, R. (1966). Electromagnetism as an aspect of geometry? Already unified field theory—The null field case. *Annals of Physics*, *36*, 147–187.

Gray, J. (1999). *The symbolic universe: Geometry and physics 1890–1930*. Oxford: Oxford University Press.

Gray, J. (2000). *The Hilbert challenge*. New York: Oxford University Press.

Guth, E. (1970). Contribution to the history of Einstein's geometry as a branch of physics. In M. Carmeli, S. Fickler, & L. Witten (Eds.), *Relativity: Proceedings of the relativity conference in the midwest, Cincinnati, Ohio, June 2–6, 1969* (pp. 161–207). New York: Plenum Press.

Hallett, M. (1994). Hilbert's axiomatic method and the laws of thought. In A. George (Ed.), *Mathematics and mind* (pp. 158–200). New York: Oxford University Press.

Hallett, M., & Majer, U. (2004). *David Hilbert's lectures on the foundations of geometry 1891–1902*. Berlin/Heidelberg/New York: Springer.

Hilbert, D. (1899). *Grundlagen der Geometrie. (Festschrift zur Einweihung des Göttinger Gauss-Weber Denkmals)*. Leipzig. Subsequent editions Leipzig: B. G. Teubner. Translation of 10th German edition by L. Unger as *Foundations of geometry.* La Salle, IL: Open Court, 1971.

Hilbert, D. (1901). Mathematische Probleme. Archiv für Mathematik und Physik, 1, 44–63 and 213–237. Reprinted in Hilbert (1935, pp. 290–325). Translation as *Hilbert's lecture at the international congress of mathematicians*. In Gray (2000, pp. 240–282).

Hilbert, D. (1915a). Die Grundlagen der Physik (Erste Mitteilung). Annotated "Erste Korrektur meiner erste Note," printer's stamp date "6 Dez. 1915." 13 pages with omissions. Translation as *The foundations of physics* (proofs of first communication). In Renn and Schemmel (2007, pp. 989–1001).

Hilbert, D. (1915b). Die Grundlagen der Physik (Erste Mitteilung). *Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematische-Phyikalische Klasse,* 395–407. Translation by D. Fine as *The foundations of physics* (first communication). In J-P Hsu & D. Fine (Eds.), *100 years of gravity and accelerated frames: The deepest insights of Einstein and Yang-Mills.* Singapore: World Scientific, 2005, pp. 120–131; translation with same title in Renn and Schemmel (2007, pp. 1003–1015).

Hilbert, D. (1916a). *Die Grundlagen der Physik. Typescript of summer semester lecture notes*. Göttingen: Bibliothek des Mathematisches Institut, Universität Göttingen 111pp.

Hilbert, D. (1916b). Das Kausalitätsprinzip in der Physik. Typescript of lectures, dated 21 and 28 November 1916 by Sauer (2001). Bibliothek des Mathematisches Institut, Universität Göttingen, 17pp.

Hilbert, D. (1916–1917). Die Grundlagen der Physik II. Typescript of winter semester lecture notes, annotated by R. Bär. Bibliothek des Mathematisches Institut, Universität Göttingen. 199pp.

Hilbert, D. (1917). Die Grundlagen der Physik (Zweite Mitteilung). *Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematische-Phyikalische Klasse,* 53–76. Translation as *The foundations of physics* (second communication). In Renn and Schemmel (2007, pp. 1017–1038).

Hilbert, D. (1918). Axiomatisches Denken*, Mathematische Annalen* 78, 405–415; reprinted in Hilbert (1935), pp. 146–156. Translation by W. Ewald as *Axiomatic thought*. In W. Ewald, From *Kant to Hilbert: A sourcebook in the foundations of mathematics*: Vol. 2. New York: Oxford University Press, pp. 1107–1115.

Hilbert, D. (1919–1920). In D. Rowe (Ed.), *Natur und mathmatisches Erkennen. Vorlesungen, gehalten 1919–1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays, 1992*. Basel–Boston–Berlin: Birkhäuser.

Hilbert, D. (1922). Neubegründung der Mathematik. Erste Mitteilung. *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, *1*, 157–177 Reprinted in Hilbert (1935, pp. 157–177). Translation by W. Ewald as *The new grounding of mathematics*. First report. In W. Ewald, *From Kant to Hilbert: A sourcebook in the foundations of mathematics*: Vol. 2. New York: Oxford University Press, pp. 1117–1134.

Hilbert, D. (1923). Die Weltgleichungen. Three lectures given in summer 1923 in Hamburg. To appear in a forthcoming volume on Hilbert's lectures on physics, published by Springer, edited by U. Majer & T. Sauer.

Hilbert, D. (1924). Die Grundlagen der Physik. *Mathematische Annalen*, *92*, 1–32 Reprinted with alternations and annotations in Hilbert (1935), pp. 258–289.

Hilbert, D. (1926). Über das Unendliche. *Mathematische Annalen*, *95*, 161–190 Translation by S. Bauer-Mengelberg as *On the infinite*. In J. van Heijenoort (Ed.), *From Frege to Gödel: A sourcebook in mathematical logic, 1879–1931*. Cambridge, MA: Harvard University Press, 1967, pp. 367–392.

Hilbert, D. (1930). Naturerkennen und Logik. *Die Naturwissenschften*, 17, 959–963; reprinted in Hilbert (1935, pp. 378–387).

Hilbert, D. (1935). *Gesammelte Abhandlungen. Bd. III*. Berlin: J. Springer.

Hoefer, C. (2000). Energy conservation in GTR. *Studies in History and Philosophy of Modern Physics*, *31*, 187–199.

Howard, D., & Norton, J. (1993). Out of the labyrinth? Einstein, Hertz, and the Göttingen answer to the hole argument. In J. Earman, M. Janssen, & J. Norton (Eds.), *The attraction of gravitation: New s tudies in the history of general relativity (Einstein Studies: Vol. 5)* (pp. 30–62). Boston/Basel/Berlin: Birkhäuser.

Janssen, M., & Renn, J. (2007). Untying the knot: How Einstein found his way back to field equations discarded in the Zurich notebook. In M. Janssen, J. D. Norton, J. Renn, T. Sauer, & J. Stachel (Eds.), *The Genesis of General Relativity: Vol. 2. Einstein's Zurich notebook: Commentary and essays* (pp. 839–925). Dordrecht: Springer.

Klein, F. (1917). Zu Hilberts erster Note über die Grundlagen der Physik. *Nachrichten der Königliches Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse.* pp. 469–482. Reprinted with additional remarks in Klein, *Gesammelte Abhandlungen,* Bd.I. Berlin: J. Springer, 1921, pp. 553–567.

Klein, F. (1927). *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert. Teil 2*. Berlin: J. Springer.

Landau, L., & Lifshitz, E. (1975). *The classical theory of fields.* Fourth rev. Translated from the Russian by M. Hamermesh. Oxford/New York: Pergamon Press.

Lorentz, H., et al. (1923). *The principle of relativity: A collection of original memoirs on the special and general theory of relativity*. London: Methuen Reprint New York: Dover, 1952.

Majer, U. (1993a). Hilberts Methode der Idealen Elemente und Kants regulativer Gebrauch der Ideen. *Kant-Studien*, *84*, 51–77.

Majer, U. (1993b). Different forms of finitism. In J. Czermak (Ed.), *Philosophie der Mathematik* (pp. 185–194). Wien: Verlag Hölder-Pichler-Tempsky.

Majer, U. (1995). Geometry, intuition and experience: From Kant to Husserl. *Erkenntnis*, *42*, 261–285.

Majer, U. (2001). The axiomatic method and the foundations of science: Historical roots of mathematical physics in Göttingen (1900–1930). In M. Rédei, & M. Stölzner (Eds.), *John von Neumann and the foundations of quantum physics* (pp. 11–33). Dordrecht: Kluwer.

Majer, U., & Sauer, T. (2004). Hilbert's 'world equations' and his vision of a unified science. In A. J. Kox, & J. Eisenstaedt (Eds.), *The universe of general relativity (Einstein Studies: Vol. 11)* (pp. 259–276). Boston/Basel/Berlin: Birkhäuser.

Mehra, J. (1974). *Einstein, Hilbert, and the theory of gravitation: Historical origins of general relativity theory*. Dordrecht: D. Reidel.

Mie, G. (1912). Grundlagen einer Theorie der Materie. *Annalen der Physik*, *37*, 511–534 Zweite Mitteilung, *Annalen der Physik* 39, 1–40.

Mie, G. (1913). Grundlagen einer Theorie der Materie. *Dritte Mitteilung, Annalen der Physik*, *40*, 1–66.

Mie, G. (1917). Die Einsteinsche Gravitationstheorie und das Problem der Materie. *Physikalische Zeitschrift*, *18*, 551–556 574–80, 596–602.

Misner, C., & Wheeler, J. (1957). Classical physics as geometry. In J. Wheeler (Ed.), *Geometrodynamics* (pp. 225–307). New York: Academic Press 1962.

Neiman, S. (1994). *The unity of reason: Rereading Kant*. New York: Oxford University Press.

Nelson, L. (1928). Kritische Philosophie und mathematische Axiomatik, *Unterrichtsblätter für Mathematik und Naturwissenschaften;* as reprinted in L. Nelson, *Beiträge zur Philosophie der Logik und Mathematik.* Frankfurt a.M.: Verlag Öffentliches Leben, 1959, pp. 89–124.

Noether, E. (1918). Invariante Variationsprobleme, Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematische-Phyikalische Klasse, pp. 235–257. Translation by M. Tavel (1971) as Noether's theorem. In Transport Theory and Statistical Physics: Vol. 1, pp. 183–207.

Norton, J. (1984). How Einstein found his field equations: 1912–1915. *Historical Studies in Physical Science*, *14*, 253–316.

Norton, J. (1993). General covariance and the foundations of general relativity. *Reports on Progress in Physics*, *56*, 791–858.

Pais, A. (1982). *'Subtle is the Lord…' The life and science of Albert Einstein*. New York: Oxford University Press.

Pauli, W., Jr. (1921). Relativitätstheorie. *Encyklopädie der mathematischen Wissenschaften,* v.19 Leipzig: B.G. Teubner. Translation as *The theory of relativity.* Oxford and New York: Pergamon Press; 1958.

Pauli, W., Jr. (1979). Dkghid. In Bd. I. A. Hermann, K. v. Meyenn, & V. Weisskopf (Eds.), *Wolfgang Pauli. Wissenschaftlicher Briefwechsel*. Heidelberg/Berlin/New York: Springer Verlag.

Peckhaus, V. (1990). *Hilbertprogramm und Kritische Philosophie*. Göttingen: Vandenhoeck und Ruprecht.

Peckhaus, V. (1994). Hilbert's axiomatic programme and philosophy. In E. Knoblauch, & D. Rowe (Eds.), *The History of Modern Mathematics*, Vol. 3 (pp. 91–112). New York: Academic Press.

Pyenson, L. (1985). Relativity in late Wilhelmian Germany: The appeal to a pre-established harmony between mathematics and physics. In Pyenson (Ed.), *The young Einstein* (pp. 137–155). Bristol: Adam Hilger Ltd.

Rainich, G. (1925). Electricity in curved space-time. *Nature*, *115*(1925), 498.

Reid, C. (1970). *Hilbert*. New York/Heidelberg/Berlin: Springer.

Renn, J., & Schemmel, M. (Eds.) (2007). *The Genesis of General Relativity: Vol. 4. Gravitation in the twilight of classical physics: The promise of mathematics*. Dordrecht: Springer.

Renn, J., & Stachel, J. (1999). *Hilbert's foundation of physics: From a theory of everything to a constituent of general relativity*. Berlin: Max-Planck-Institut für Wissenschaftsgeschichte Preprint 118. Reprinted in Renn & Schemmel (Eds.), 2007, pp. 858–973.

Rowe, D. (1999). The Göttingen response to general relativity and Emmy Noether's theorems. In J. Gray (Ed.), *The symbolic universe: Geometry and physics 1890–1930* (pp. 189–233). New York: Oxford University Press.

Rowe, D. (2001). Einstein meets Hilbert: At the crossroads of mathematics and physics. *Physics in Perspective*, *3*, 379–424.

Rowe, D. (2002). Einstein's gravitational field equations and the Bianchi identities. *The Mathematical Intelligencer*, *24*, 57–66.

Rowe, D. (2003). Hermann Weyl, the reluctant revolutionary. *The Mathematical Intelligencer*, *25*, 61–70.

Ryckman, T. (2005). *The reign of relativity: Philosophy in physics 1915–1925. (Oxford studies in philosophy of science)*. New York: Oxford University Press.

Sauer, T. (1999). The relativity of discovery: Hilbert's first note on the foundations of physics. *Archive for History of Exact Sciences*, *53*, 529–575.

Sauer, T. (2001). The relativity of elaboration: Hilbert's second note on the foundations of physics. Ms. dated 10 September 2001, 14pp.

Sauer, T. (2002). Hopes and disappointments in Hilbert's axiomatic 'foundations of physics'. In M. Heidelberger, & F. Stadler (Eds.), *History of philosophy of science: New trends and perspectives* (pp. 225–237). Dordrecht: Kluwer.

Sauer, T. (2005). Einstein equations and Hilbert action: What is missing on page 8 of the proofs for Hilbert's first communication on the foundations of physics? *Archive for History of Exact Sciences*, *59*, 577–590.

Stachel, J. (1992). The Cauchy problem in general relativity—The early years. In J. Eisenstaedt, & A. J. Kox (Eds.), *Studies in the history of general relativity. (Einstein Studies: Vol. 3)* (pp. 407–418). Boston/Basel/Berlin: Birkhäuser.

Stachel, J. (1993). The meaning of general covariance: The hole story. In J. Earman, et al. (Eds.), *Philosophical problems of the internal and external world: Essays on the philosophy of Adolf Grünbaum* (pp. 129–160). Pittsburgh: University of Pittsburgh Press.

Stachel, J. (1999). New light on the Einstein–Hilbert priority question. *Journal of Astrophysics and Astronomy*, 20, 91–101; as reprinted in Stachel (2002, pp. 353–364).

Stachel, J. (2002). *Einstein from 'B' to 'Z'*. Boston/Basel/Berlin: Birkhäuser (Einstein Studies: Vol. 9).

Torretti, R. (1978). *Philosophy of geometry from Riemann to Poincaré*. Dordrecht: D. Reidel.

Trautman, A. (1962). Conservation laws in general relativity. In L. Witten (Ed.), *Gravitation: An introduction to current research* (pp. 169–198). New York: Wiley.

Vermeil, H. (1917). Notiz zur das mittlere Krümmungsmass einer n-fach ausgedehnten Riemann'schen Mannigfaltigkeit. *Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse,* pp. 334–344.

Vizgin, V. (1994). *Unified field theories in the first third of the 20th century.* Translated from the Russian by J. Barbour. Boston/Basel/Berlin: Birkhäuser.

Vizgin, V. (2001). On the discovery of the gravitational field equations by Einstein and Hilbert: New materials. *Physics-Uspekhi*, *44*(12), 1283–1298.

Von Laue, M. (1923). *Die Relativitätstheorie, Zweiter Band: Die allgemeine Relativitätstheorie und Einsteins Lehre von der Schwerkraft*. Braunschweig: Friedrich Vieweg & Sohn.

Weyl, H. (1918a). *Raum-Zeit-Materie*. Berlin: J. Springer.
Weyl, H. (1918b). Reine Infinitesimalgeometrie. *Mathematische Zeitschrift*, *2*, 384–411 Reprinted in Weyl, *Gesammelte Abhandlungen* Bd. 2. Berlin: Springer; 1968, pp. 1–28.
Weyl, H. (1921). *Raum-Zeit-Materie. Vierte Auflage*. Berlin: J. Springer Translation by H. Brose as *Space-time-matter*. Reprint New York: Dover, 1953.
Wightman, A. (1976). Hilbert's sixth problem: Mathematical treatment of the axioms of physics. In *Proceedings of symposia in pure mathematics*. Vol. 28, pp. 147–240.