

# Hilbert's 'Foundations of Physics': Gravitation and Electromagnetism Within the Axiomatic Method

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)

K.A. Brading<sup>†</sup> and T.A. Ryckman<sup>‡</sup>

## *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 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 contribution 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 contributions 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 contribution is a fine and coherent piece of work within this framework, whose principal aim is to address the apparent tension between general invariance and causality (in the precise sense of Cauchy determination), pinpointed in Theorem I of the first contribution. This is not the same problem as that found in Einstein’s “hole argument” – something that, we argue, never confused Hilbert.

- 1 Introduction
- 2 The essential context: Hilbert’s axiomatic method and Kantian epistemology
- 3 Hilbert’s First Communication on ‘The Foundations of Physics’ (published version)
  - 3.1 Hilbert’s aim
  - 3.2 Schematic outline
  - 3.3 The connection between gravitation and electromagnetism

---

<sup>†</sup> Department of Philosophy, 100 Malloy Hall, University of Notre Dame, IN 46556. kbrading@nd.edu.

<sup>‡</sup> Department of Philosophy, Stanford University, Stanford, CA 94305. tryckman@stanford.edu.

4	Comparison of the December Proofs with the published version of the First Communication
4.1	Einstein's field equations
4.2	Hilbert's target: the 'problem of causality'
5	The First Communication and the Axiomatic Method
6	Hilbert's Second Communication on 'The Foundation of Physics'
6.1	Secondary literature on the Second Communication
6.2	Outline of Hilbert's revision of Kant
6.3	Outline of Hilbert's Second Communication
6.4	Preliminaries
6.5	The problem of causal order
6.6	The problem of univocal determination
6.7	Physics as a 4-Dimensional Pseudogeometry
6.8	Conclusion concerning the Second Communication
7	Hilbert and the 'hole argument'
8	Hilbert's revision of Kant in the light of general invariance
9	Conclusions
	Acknowledgements
	References
	Appendix 1: A sketch of the Cauchy problem in general relativity
	Appendix 2: Chronology

## 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' and 'Second Communication', two papers (or 'notes', as they are widely known) eventually appeared in the *Nachrichten* of the Academy.<sup>1</sup> 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 dated 6 December 1915 (henceforth the 'December Proofs').<sup>2</sup> With the exception of Renn and Stachel (1999), the Second Communication has not been given the same detailed reconsideration. However, Renn and Stachel's analysis of the Second Communication seriously misrepresents its aims, content, and significance, and also its links with 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.

---

<sup>1</sup> Hilbert (1915b), (1917). English translations of these papers, as well as of Hilbert (1915a), are available in Renn and Schemmel, eds., (2007). Unless otherwise noted, all translations in this paper are our own.

<sup>2</sup> Hilbert (1915a), and Corry, Renn and Stachel (1997); see e.g., Rowe (1999) and (2001), Renn and Stachel (1999), Stachel (1999), Sauer (1999), Vizgin (2001), Corry (2004), Sauer (2005).

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.<sup>3</sup> 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' 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.<sup>4</sup> Correspondingly, some historians of general relativity have concluded that Hilbert 'attached a kind of metaphysical significance to variational methods',<sup>5</sup> 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 physics.<sup>6</sup> 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<sup>7</sup> or simply led to hopelessly failed physics.<sup>8</sup>

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

---

<sup>3</sup> Corry (2004), treating Hilbert's notes as part of his program for the axiomatization of physics, is a welcome exception.

<sup>4</sup> Renn and Stachel (1999), 1; Reid (1970), 127.

<sup>5</sup> Rowe (1999), 201.

<sup>6</sup> Pyenson (1985). 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 (1919-20), 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*).

<sup>7</sup> 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 1).

<sup>8</sup> Stachel (1992); Renn and Stachel (1999), 77; 81-83.

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 towards 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.<sup>9</sup> Unlike Einstein’s largely informal and heuristic extraction from the clutches of 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 once seem assured).’<sup>10</sup> 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 objectivity* as this is understood in a broadly transcendental idealist sense, most prominently displayed in his subsequent remarks regarding the axiom of general

---

<sup>9</sup> See Norton (1984, pp. 286-291, and 1993, sections 1-3), Stachel (1993), and Ryckman (2005, Section 2.2.2), for presentation and discussion of the ‘hole argument’, and for additional references.

<sup>10</sup> ‘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 on p. 566, fn 8, added to the 1921 reprint of Klein (1917).

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 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.<sup>11</sup> 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.<sup>12</sup> 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.<sup>13</sup>

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.<sup>14</sup> 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

---

<sup>11</sup> Einstein to H. Weyl, 23 Nov. 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 (1998A), 366.

<sup>12</sup> E.g., Guth (1970), 84; Mehra (1974), 26, 72, n. 145; Wightman (1976), 153; Pais (1982), 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.

<sup>13</sup> Hallett (1994), 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.'

<sup>14</sup> Hallett and Majer (2004), 66.

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 geometry subjected to the axiomatic treatment. In point of fact, Hilbert regarded the axiomatization as 'the logical analysis of our spatial intuition'.<sup>15</sup> 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 in 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'.<sup>16</sup> As the axiomatic method is characterized in his Göttingen Winter Semester lectures in 1922/23, analysis next determines the concepts under which such given facts can be classified and arranged, and then 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*)

---

<sup>15</sup> '*Die bezeichnete Aufgabe läuft auf die logische Analyse unserer räumlichen Anschauung hinaus.*' Hilbert (1899), 3.

<sup>16</sup> Hilbert (1922), 161; Engl. trans., 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.

of the (axiomatic) framework and the real things of the actual world, and then to have carried this through.<sup>17</sup>

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.<sup>18</sup> In each case, the aim is not, at least in the first instance, the discovery or recognition of *new* laws or principles, but the conceptual and logical clarification or reconstruction of known ones.<sup>19</sup> Finally, and as its culmination, the axiomatic method is concerned to demonstrate that the axioms of the theory thus selected possess the three meta-logical properties or relations of mutual consistency, independence, and completeness.<sup>20</sup> 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.<sup>21</sup>

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.<sup>22</sup> Yet the axiomatic method requires 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’.<sup>23</sup> This emphasis on

---

<sup>17</sup> Hilbert Winter Semester lectures 1922/23 *Wissen und mathematisches Denken*. Ausgearbeitet von Wilhelm Ackermann. Mathematisches Institut Göttingen. Published in a limited edition, Göttingen, 1988; as translated in Hallett (1994), 167.

<sup>18</sup> Nelson (1928) is 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 translated in Peckhaus (1994), 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’.

<sup>19</sup> Cf. Majer (2001), 19.

<sup>20</sup> 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), 59.

<sup>21</sup> Hilbert (1918), 407 (Engl. transl., 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.’

<sup>22</sup> See, e.g., the discussion in Corry (2004), 112-114.

<sup>23</sup> Lecturing in 1918 on the ‘Basic Ideas of General Relativity’, Hilbert stressed that the new conceptions of space, time and motion of Einstein’s theory were still compatible with ‘the traditional intuition’ of

the compatibility between the different sources of knowledge is crucial to 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 axiomatic method, in virtue of their ideality, and so severance from experience and intuition, axioms can play at best a hypothetical [deletion] 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'.<sup>24</sup> He concluded that what remains of Kant's synthetic *a priori* is just this '*a priori* intuitive 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 *a priori* intuitive point of view (*jene anschauliche Einstellung a priori*) and thereby of investigating the conditions of the possibility of all conceptual knowledge and of all experience.<sup>25</sup>

---

'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 space-time 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), 274.

<sup>24</sup> Hilbert (1930), 961.

<sup>25</sup> 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 '*a priori* intuitive 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 1922/23 lectures. Ausgearbeitet von W. Ackermann. Mathematisches Institut Göttingen. Published in a limited edition, Göttingen, 1988. As cited and translated in Corry (2004), 429.

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.<sup>26</sup> However, with the discovery of the December Proofs in 1996 it was learned that the version submitted on 20 November differs considerably from that appearing in the published *Nachrichten* of 31 March 1916.<sup>27</sup> 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,<sup>28</sup> our purpose is rather different. Our interest lies in comparing the content of the First Communication, and especially what was cut from the December Proofs, with what was published as the content of the Second Communication. This enables us to see that the Second Communication treats in detail an issue raised in December 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.<sup>29</sup>

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 announces that his aim is to set up '*in the sense of the axiomatic method*'<sup>30</sup> 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

---

<sup>26</sup> Rowe (2001), 418.

<sup>27</sup> Hilbert was sending offprints to colleagues in mid-February 1916; see Sauer (1999), 543, n.74.

<sup>28</sup> Especially in Sauer (1999), and (2005), Renn and Stachel (1999), and Vizgin (2001).

<sup>29</sup> Hilbert (1915b), 395; (1917), 63-64; (1923).

<sup>30</sup> Hilbert (1915a), 1; (1915b), 395: "*im Sinne der axiomatischen Methode*". Our emphasis added.

both Einstein's theory of gravitation and Gustav Mie's theory of matter.<sup>31</sup> These two theories were, in 1915, clear candidates for the fundamental theories of physics.<sup>32</sup> 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 particular 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 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'  $H$ ,<sup>33</sup> depending upon 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:

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

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

---

<sup>31</sup> See in particular the discussion of Mie's theory in Corry (2004), 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).

<sup>32</sup> Einstein's antipathy toward Mie's theory of matter is well-known. It also anteceded Hilbert's reformulation of the theory. Referring to the Mie theory in a letter to Erwin Finlay Freudlich of mid-August 1913 he termed it 'fantastic' remarking, 'in my opinion, it has only a vanishing small inner probability' (Einstein, 1993, 550). However, Sauer (2002, 231) notes (in response to claims by Renn and Stachel as to the implausibility in 1915 of the Mie theory): "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 treatise on general relativity contains a 5 page section (§29) on Mie's theory of matter *in the context of Einstein's theory*.

<sup>33</sup> The term 'world function', appearing in Mie (1912-1913), while reminiscent of Minkowski, was used by Hilbert already in 1905 lectures on Newtonian continuum mechanics; see Corry (2004), 152.

Hilbert's use of the term 'world parameters' in place of the standard locution 'space-time coordinates' is instructive. As expressly stated in his Second Communication, and as Mie noted that same year,<sup>34</sup> it is intended to highlight the analogy Hilbert sought to draw between the arbitrariness of parameter representations of curves in the calculus of variations, and the arbitrariness of coordinates on a space-time 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 space-time coordinates were already in place. Similarly, in both versions 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.<sup>35</sup>

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 'world parameters' has been dropped while the sentence has been reformulated explicitly in terms of the physical meaninglessness of space-time 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  $q_s$ .<sup>36</sup>

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'.<sup>37</sup>

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

---

<sup>34</sup> Hilbert (1917), 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), 599, also stressed this analogy.

<sup>35</sup> Hilbert (1915a), 2; (1915b), 396.

<sup>36</sup> Hilbert (1924), 4.

<sup>37</sup> Corry (2004), 401.

independent combinations of the  $n$  equations and their total derivatives are always identically satisfied.<sup>38</sup>

One of Hilbert's principal claims is that, as a consequence of Theorem I, electromagnetic phenomena may be regarded as consequences of the gravitational. 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 (using Hilbert's numbering), on the one hand, ten equations for the gravitational potentials,

$$(4) \quad \frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g}H}{g_{kl}^{\mu\nu}} = 0 \quad , \text{ or, in Hilbert's abbreviation,}$$

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

while, on the other, differentiation of  $H$  with respect to the electromagnetic potentials  $q_s$  and their first derivatives yields four equations,<sup>39</sup>

$$(5) \quad \frac{\partial \sqrt{g}H}{\partial q_h} - \sum_\sigma \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial q_{hk}} = 0 \quad ,$$

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

The fourteen equations (4) and (5) 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.

---

<sup>38</sup> Hilbert (1915a), 2-3; (1915b), 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), 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.

<sup>39</sup> The form of equations (4) and (5) 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 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'.

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 (5) are a consequence of the ten equations (4), such that, '*in the sense indicated (in dem bezeichneten Sinne), electrodynamic phenomena are effects of gravitation*'.<sup>40</sup> 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 and by Corry.<sup>41</sup> 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 space-time geometry, the curvature of space-time as expressed by the Riemann tensor.<sup>42</sup> 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.<sup>43</sup> 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 consequence of the structure of the gravitational part of the theory.<sup>44</sup> 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 bellwether 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:

---

<sup>40</sup> Hilbert (1915a), 3; (1915b), 397.

<sup>41</sup> Renn and Stachel (1999), 36-41; Corry (2004), 336-7.

<sup>42</sup> Misner and Wheeler (1957), Geroch (1966). These papers follow up earlier results of Rainich. See, e.g., Rainich (1925), 498: 'It is often thought that the theory of curved space-time (general relativity theory) accounts for gravitation but *does not account for the electromagnetic phenomena*. This is not so.'

<sup>43</sup> Hilbert's treatment of energy is discussed in detail in Sauer (1999), 554-557.

<sup>44</sup> Hilbert's considerations on the tension between general covariance and causality (see section 4.2, below) are the other central outcome.

$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0$ ,<sup>45</sup> where the first term is re-written, in the published version but not in the Proofs, with the crucial trace term (see section 4.1 below),

$[\sqrt{g}K]_{\mu\nu} = \sqrt{g} (K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu})$ . Now  $L$  is a general invariant that is assumed to depend *only* on the  $g_{\mu\nu}$ , the  $q_s$ , and the first derivatives  $\frac{\partial q_s}{\partial w^i}$ . 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

$$(16) \quad \frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0.$$
<sup>46</sup>

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  $q_s$

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

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.<sup>47</sup> 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.*<sup>48</sup>

The assumption that nothing else beyond the  $g_{\mu\nu}$  (but no derivatives of the metric), the  $q_s$ ,

and the so-constrained first derivatives  $\frac{\partial q_s}{\partial w^i}$  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  $L$ .<sup>49</sup> It also has consequences for the interpretation of the energy-momentum tensor  $T_{\mu\nu}$  in Hilbert’s theory. Since all non-gravitational energy/matter is contained in  $L$ , it is entirely sufficient for forming  $T_{\mu\nu}$ , *i.e.*,

<sup>45</sup> Hilbert (1915a), 11 eq. (26); Hilbert (1915b), 404, eq. (21).

<sup>46</sup> Hilbert (1915a), 10 eq. (22); Hilbert (1915b), 403, eq. (16).

<sup>47</sup> See Born (1914), 28.

<sup>48</sup> Hilbert (1915b), 403; at the corresponding place in Hilbert (1915a), 10, the bracketed expression does not appear.

<sup>49</sup> Born (1914), 24-25.

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

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

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

do not have the same interpretation, because Hilbert assumed a particular hypothesis about the electromagnetic constitution of all matter.<sup>50</sup> 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.<sup>51</sup> 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.*<sup>52</sup>

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 towards 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 4 linearly independent identities containing the Euler derivative associated with the electromagnetic equations,

$$(27) \quad \sum_m (M_{m\nu} [\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m} [\sqrt{g}L]_m) = 0,$$

where we again recall that

<sup>50</sup> Earman and Glymour (1978), 303; Sauer (1999), 564.

<sup>51</sup> For details, see Sauer (1999), 555.

<sup>52</sup> Hilbert (1915a), 10; (1915b), 404. Note that *already in the December Proofs*, Hilbert referred here to 'Einstein's general theory of relativity' (*der Einsteinschen allgemeinen Relativitätstheorie*), explicitly according due credit to Einstein.

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

are the abbreviated Lagrangian form of the electromagnetic field equations.<sup>53</sup> 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 made following his statement of Theorem I, Hilbert claimed, again in italic type for emphasis:

*This is the exact mathematical expression of the above generally stated assertion concerning the character of electrodynamics as an accompanying phenomenon (Folgeerscheinung) of gravitation.*

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.<sup>54</sup> 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 1,' a special case of Noether's second theorem, made it possible to regard the equations of electrodynamics as consequences of the gravitational field equations.<sup>55</sup>

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 stated more broadly for any generally invariant theory, Theorem I states 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 (although this is not always explicitly stated).<sup>56</sup> 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

<sup>53</sup> Hilbert (1915a), 12 eq. (32); (1915b), 406 eq. (27). See also Sauer (1999), 556-557.

<sup>54</sup> Rowe (2001), 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.

<sup>55</sup> Vizgin (1994), 58-59. However, in view of the just given derivation, we do not agree with Vizgin's subsequent analysis (61 ff) that Hilbert's choice of just *which* 4 equations followed from the other  $n-4$  (according to Theorem I) was arbitrary.

<sup>56</sup> Noether (1918).

of material phenomena completely, if the latter is capable of being characterized by 4 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 of the December Proofs on which appears Theorem I.<sup>57</sup>

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 assessment of Rowe (2001, 412) 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...’. 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.<sup>58</sup> 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.<sup>59</sup>

#### **4 Comparison of the December Proofs with the published version of the First Communication**

Felix Klein is reported to have commented upon the ‘completely disordered’ character of Hilbert’s First Communication, remarking that it was ‘evidently a product of great exertion and excitement’.<sup>60</sup> 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.<sup>61</sup> 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) 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

---

<sup>57</sup> Page 151 of Parret and Jeffrey translation of Einstein, 1916. In the original text, (1916, 810; reprinted (1996), 325), Einstein referred to page 3 of Hilbert (1915b); as Sauer (1999, 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 p. 395 (the first page) of the published version. In fact, it is to p. 397 of that version, containing Theorem I. See Janssen and Renn (2007) on Einstein’s prior use of variational techniques and his recognition that energy conservation is connected with the four identities.

<sup>58</sup> E.g., Trautman (1962), 334.

<sup>59</sup> Pauli (1921), §55.

<sup>60</sup> Klein, Göttingen Lecture Notes, 10 Dec. 1920 [Klein *Nachlaß* XXII C, p.18], as cited and translated in Rowe (2002), 61.

<sup>61</sup> See Klein (1917) for his ‘simplification’ (*Vereinfachung*) of Hilbert’s treatment; for discussion, see Rowe (1999), 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’s to Einstein of 27 May 1916. Einstein (1998A, 289-295.)

structure of the gravitational part of the theory (see section 3.3), and (2) an 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 December 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 December 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 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 (though with the qualifications of section 3.3 above) generally covariant gravitational field equations as those of Einstein. However, in the December 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 gravitational equations occur explicitly in Einstein's 25 November presentation to the Prussian Academy that had appeared 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 December Proofs provide evidence of 'nostrification' by Hilbert of Einstein's final results,<sup>62</sup> and 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 essentially combined Einstein's earlier *non-covariant Entwurf*

---

<sup>62</sup> Famously, in a letter of 26 November 1915, Einstein complained to his friend Heinrich Zangger in Zürich, that "only one colleague has actually understood" his new theory and that person had "sought to 'nostrify' it (Abraham's expression) in a clever way" – a clear reference to Hilbert. (Einstein (1998), 204-05). Corry (2004), 99 and 419-22, observing that the term is "ambivalent and subtle", notes 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.

theory of gravitation with Mie's theory.<sup>63</sup> 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 central to 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 December Proofs. In the published version, Hilbert identifies his world function as composed of two additive parts,

$$H = K + L,$$

the gravitational and matter components.<sup>64</sup> In the 1924 republication of this note, this supposition is stated as an additional Axiom (III) further specifying the world function. However, in the Proofs, some text is missing containing the equation numbered (17). The above specification occurs at the corresponding place in the published version, and *so was almost certainly contained in the December Proofs*.<sup>65</sup> In both the Proofs and the published version,  $K$  is identified, without proof, as the only scalar invariant depending only on the  $g_{\mu\nu}$  and its first and second derivatives. 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.<sup>66</sup> The gravitational field equations then appear several pages later (in the numbering of the Proofs) in the form of Lagrangian derivatives:

$$(26) \quad \left[ \sqrt{g} K \right]_{\mu\nu} + \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} = 0.$$

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*.<sup>67</sup>

In the published version, but *not* in the Proofs, Hilbert noted that 'it follows easily without calculation' that

---

<sup>63</sup> Renn and Stachel (1999), 35; Stachel (1999), 359. We agree, according, with Sauer (1999), 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.'

<sup>64</sup> See note 38 above.

<sup>65</sup> See Sauer (2005).

<sup>66</sup> Hilbert (1915a), 8; (1915b), 402. Hilbert failed to state that 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 in Landau and Lifshitz (1975), 268. Rowe (2001), 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 (Engl. trans. 315-317). Both Rowe and Weyl trace the result to a paper by Felix Klein's assistant Hermann Vermeil (1917). On the latter's contribution, see Sauer (2005).

<sup>67</sup> This has been particularly emphasized by Sauer (1999), 547.

$$\left[ \sqrt{g} K \right]_{\mu\nu} = \sqrt{g} \left( K_{\mu\nu} - \frac{1}{2} K g_{\mu\nu} \right).^{68}$$

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 December Proofs was devoted to constructing the non-covariant energy theorem, Corry, Renn, and Stachel (1997), 1272, allege that ‘knowledge of Einstein’s result may have been crucial to Hilbert’s introduction of the trace term into his field equations.’ 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.<sup>69</sup> In fact, Hilbert presented an explicit calculation only in the 1924 republication of his two notes in *Mathematische Annalen*.<sup>70</sup> 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 project. Indeed, we think that the weight of evidence agrees with Corry’s recent 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.<sup>71</sup>

Nonetheless, even though the *explicit* form of the field equations does not appear in the December Proofs, nevertheless – as already noted – Hilbert’s gravitational field equations as given *implicitly* in terms of his Lagrangian *are* generally covariant already in the December Proofs.

## 4.2 Hilbert’s target: the ‘problem of causality’

Both the December 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

---

<sup>68</sup> Hilbert (1915b), 405.

<sup>69</sup> Sauer (1999), 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.”

<sup>70</sup> See Renn and Stachel (1999), 64-65.

<sup>71</sup> Corry (2004), 403.

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 December Proofs, but not in the published version, Hilbert explicitly spells out the implications of Theorem I for his system of fundamental equations of physics.

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  $g_{\mu\nu}, q_s$ , 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 (4) and (5) is essential.<sup>72</sup>

Thus, independent of the physical validity of his system of fundamental equations, for which he adduced no evidence whatsoever, Hilbert clearly underscored 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 a seeming conflict 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.<sup>73</sup> Theorem I suggests that it is a property of any such theory that it does not.<sup>74</sup> For a system of second order partial differential equations, the Cauchy problem is that of showing 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 away from that region (the region's 'domain of dependence'). For field theories formulated in space-time, the initial data are formulated on a given space-like hypersurface  $\Sigma$ , 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 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 space-time theory by the freedom to make arbitrary coordinate transformations (equivalently, diffeomorphic point transformations) of solutions to the field equations. Formulated for a generally invariant Lagrangian by Hilbert's Theorem I, this is the fact that not all of 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  $n-4$  space-time equations. Thus, theorem I is *a precise mathematical statement of the tension between the*

---

<sup>72</sup> Hilbert (1915a), 3-4.

<sup>73</sup> See the Appendix for a sketch of the Cauchy problem in general relativity.

<sup>74</sup> We discuss Hilbert's analysis in relation to Einstein's 'hole argument' in section 7, below.

*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 December Proofs of an ‘energy form’

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

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  $\sqrt{g}P_g H$ , 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”,

$$(15) \quad \sum_l \frac{\partial e_s^l}{\partial w_l} = 0 ,$$

and then requiring that, *for special space-time coordinates  $w_\mu$  adapted to this ‘energy theorem’*, the theorem holds.<sup>75</sup> The ‘energy theorem’ is not generally covariant, and

---

<sup>75</sup> We re-write the energy form as  $E = \sum_s e_s p^s + \sum_{s,l} \frac{\partial}{\partial w_l} (e_s^l p^s) + \sum_{s,l} p^s \frac{\partial e_s^l}{\partial w_l}$ , and discard the divergence

term to arrive at  $E = \sum_s e_s p^s + \sum_{s,l} p^s \frac{\partial e_s^l}{\partial w_l}$ . Hilbert then notes that the ‘energy theorem’ holds iff  $e_s = 0$ ,

which in turn holds iff  $\frac{d^{(g)}\sqrt{g}H}{dw_s} = 0$ . Notice that it is the *coordinate* derivative of  $\sqrt{g}H$  that vanishes

when the ‘energy theorem’ holds, not the *covariant* derivative.

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

**AXIOM III:** ('The Axiom of Space and Time'): 'The space-time coordinates are such particular world parameters for which the energy theorem (15) is valid.'

Elucidating this result, Hilbert clarified the main point, that these 4 non-covariant equations complete the system of fundamental equations of physics:

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.<sup>76</sup>

The idea that satisfaction of energy conservation (the energy theorem (15)) requires 4 non-covariant equations is almost certainly taken from the Einstein-Grossmann *Entwurf* theory of 1913-1914,<sup>77</sup> 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 reject the view of Renn and Stachel, who regard Hilbert's energy construction as his 'Proofs argument, based on causality, against general covariance'.<sup>78</sup> 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).<sup>79, 80</sup> We return to this in our discussion of the hole argument, below (section 7).

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 the 'final' version of Einstein's theory, this turned out to be the wrong approach for solving the tension between general covariance and Cauchy-determination. Hilbert dropped it altogether, modifying and truncating his treatment of energy. There – consistent with the implicitly generally covariant energy in Einstein's

---

<sup>76</sup> Hilbert (1915a), 7.

<sup>77</sup> Renn and Stachel (1999), 32, note Einstein's conviction "(e)ven before Einstein developed the hole argument", that energy-momentum conservation requires such a restriction.

<sup>78</sup> Renn and Stachel (1999), 73.

<sup>79</sup> This is also pointed out by Sauer (2005, n. 5) who writes '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.

<sup>80</sup> As Sauer (2005, n. 5) observes, Renn *et al.* (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 December Proofs differ from the modern understanding of coordinate conditions since *all applications* of the field equations require their satisfaction.

final 25 November presentation to the Berlin Academy<sup>81</sup> – 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.

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.<sup>82</sup>

## 5 The First Communication and the Axiomatic Method

We recall that the task of the axiomatization of physics was the 6<sup>th</sup> 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 assumptions, even while his theory is under development. For the consistency of these new hypotheses, while he depends, with respect to the compatibility of the new hypotheses with the old axioms, solely upon these experiments or upon 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.<sup>83</sup>

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.<sup>84</sup> For our purposes, there are three items of interest in this passage.

---

<sup>81</sup> Einstein (1915).

<sup>82</sup> 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 Hofer 2000).

<sup>83</sup> Hilbert (1900); Engl. transl., 257-258.

<sup>84</sup> 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.

- As noted 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 actuality’, 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.
- 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 60<sup>th</sup> 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.<sup>85</sup>

With regard to the Mie theory, as both Einstein and Hilbert were aware in 1915, Einstein’s gravitational theory, though in principle capable of encompassing all matter fields into space-time 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.<sup>86</sup> The axiomatic method was 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 included all known physical interactions. The principal attraction of the Mie theory to Hilbert was that, coupled with Einstein’s theory of gravitation, it enabled a hypothetical axiomatic completion of physics that could be studied by drawing consequences from the amalgamation of the two theories. In this

---

<sup>85</sup> Born (1922), 90-91.

<sup>86</sup> Einstein (1936), 335.

regard, Hilbert's 'theory' is a canonical illustration of the mode of investigation of the 'axiomatic method', in Hilbert's own most precise characterization of that method, as the 'mapping' (*Abbildung*) of a 'domain of knowledge' (*Wissensgebiet*) onto a

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 (*Wirklichkeit*). 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.<sup>87</sup>

Hilbert was familiar with the Mie theory at least since the discussion of it in Göttingen Mathematical Society in December, 1912 and again in December, 1913, when Born had set it into a more canonical mathematical form.<sup>88</sup> Certainly, the fact that Mie had sought to derive field equations of a generalized Maxwellian electrodynamics from the axiom of a Lorentz (orthogonally) invariant 'world function' appearing as a variational principle, fitted very naturally into Hilbert's axiomatic approach. Coupled with Einstein's theory of gravitation under a common variational principle, it enabled a hypothetical axiomatic completion of physics that could be studied by drawing consequences from the amalgamation of the two theories.

But the Mie theory also presented an eminently suitable candidate for the attention of the 'axiomatic method' for a number of other reasons that merit illumination, reasons not 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 SS 1913 and would again in SS 1917.<sup>89</sup> 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 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 laws, the Mie theory was clearly to be preferred to the electron theory.

Second, the Mie world function yielded four electro-dynamical equations for the four unknown electro-dynamic potentials. From given boundary and initial conditions,

---

<sup>87</sup> Hilbert's WS 1921-22 Lectures on the '*Grundlagen der Mathematik*', cited and translated in Hallett (1994), 167-168.

<sup>88</sup> Corry (1999), 176.

<sup>89</sup> Corry (1999), 174, 183. Corry (2004), 271 points out 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."

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.<sup>90</sup> 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,<sup>91</sup> he found Mie’s theory suitable to be incorporated 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.<sup>92</sup> However, both of these were withdrawn before publication, and no version of either has apparently survived. 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.<sup>93</sup> 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.<sup>94</sup> Despite period recognition by Weyl, Pauli and von Laue, this paper has been treated as largely independent of Hilbert’s First Contribution, and in any case, has been widely disregarded by historians, physicists and philosophers alike. We claim that the December 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.

---

<sup>90</sup> As a referee has reminded us, within the broad framework of Mie’s theory, one might hope to find a matter representation based on generalized Maxwell equations following from a Lagrangian containing only gauge invariant terms.

<sup>91</sup> Hilbert (1916a), 110.

<sup>92</sup> Sauer (1999), 560, note 129.

<sup>93</sup> Sauer (2001), 3.

<sup>94</sup> According to our reconstruction, this third submission must have an entirely different content from the previous (and non-extant) versions.

In the light of Einstein's General Theory of Relativity, Hilbert's resolution of the twin-faceted problem of causality as it appears 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 philosophical 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 review the existing 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, 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 originally cited this paper, but not the First Communication, in the proofs to the first edition of his classic *Raum-Zeit-Materie* (1918), much to Hilbert's chagrin when he read them.<sup>95</sup> Pauli, who also 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).<sup>96</sup> 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. On the other hand, Sauer (1999, 557 and n. 120, 560 and n. 129; 2001, 3) observes that a paper of this title had been withdrawn from publication twice by Hilbert, and in a brief summary, emphasizes that the paper does not provide what had been promised at the conclusion of the first contribution, *viz.*, the electromagnetic part of a Lagrangian world function that would allow derivation of the electron, and so explain 'the most intimate and up to now hidden processes within the atom'.

In our opinion, the December Proofs of the First Communication offer a crucial insight into the main theme of the Second Communication, providing the basis for a significant reinterpretation of the paper. It might seem curious that such a reconsideration has not been given before. However, the main motivation for reconsidering the First Communication in the light of the December Proofs has been the 'priority issue', concerning whether Einstein or Hilbert first arrived at the Einstein field equations. Given

<sup>95</sup> See the draft letter of Hilbert to Weyl dated 22 April 1918, as translated in Rowe (2003), 66.

<sup>96</sup> Pauli (1921), §22. Rowe (2001), 408, notes that Pauli's in-print recognition of Hilbert as co-discoverer 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), 27, 31.

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'.

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' (77). To be sure, much of Hilbert's paper is concerned with integrating his field equations in regions where the electromagnetic field disappears, 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* GR, 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 1916/17 Winter Semester lectures 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. (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 December 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, the new generally covariant physics requires that we modify Kant. We discuss this in detail in section 8, below, but it is worthwhile

indicating the main features of this revision now, before turning to our discussion of the details of the Second Communication.

The crux of Hilbert's suggested revision, 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 being represented in space and time and subject to the requirement that causes precede their effects remains a condition for being a possible object of experience, this is no longer a necessary criterion for being 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 a regulative constraint guiding the search for fundamental physical laws. As we shall discuss in section 8, below, Hilbert views this as a significant step forwards in removing the contributions of human subjectivity from the conceptual structure of physics.

In short, the causal principle (in distinction from the mathematical requirement of univocal determination) is an *anthropomorphic* condition imposed by the structure of the human mind on 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 some '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 6.5), and the problem of univocal determination (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 space-like and time-like coordinate axes and can never reverse the temporal order of cause and effect. Next, he pointed to the consequent need to reformulate the causality principle within 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 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 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 4-dimensional pseudo-geometry. The geometry is ‘pseudo’, Hilbert explains,<sup>97</sup> 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*)
- 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.

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 also

wrote down the Monge ordinary differential equation,  $g_{\mu\nu} \frac{dx_\mu}{dp} \frac{dx_\nu}{dp} = 0$  and the Hamilton-

Jacobi partial differential equation  $g^{\mu\nu} \frac{\partial f}{\partial x_\mu} \frac{\partial f}{\partial x_\nu} = 0$  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 world lines emanating from a world point lie inside its conoid, their boundary.<sup>98</sup>

Thirdly, it is significant that in his discussion of ‘measure threads’ and ‘light clocks’ Hilbert does not, as Einstein will several years later,<sup>99</sup> posit ‘practically rigid rods’ and ‘perfect clocks’ as presuppositions of the applicability of the pseudo-geometry

<sup>97</sup> The term ‘pseudo-geometry’, apparently derived from the term ‘pseudo-Euclidean’ applied to the metric of Minkowski space-time, appears to have been local Göttingen mathematical vernacular.

<sup>98</sup> Renn and Stachel (1999), 79; see also Klein (1927), 107.

<sup>99</sup> Einstein (1921).

to the actual world.<sup>100</sup> 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 (4) and (5) of his first contribution). As a result, Hilbert regarded measure threads and light clocks as 'instruments of invariant character' in whose terms 'every physical fact must ultimately (in principle) be ascertained'.<sup>101</sup> Presumably, Hilbert meant by this that they can be employed *everywhere* and *everywhen* by any observer, and so provide the empirical basis for all of the hypothetical total 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 two-fold. 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 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 contribution has turned out to be incorrect.

## 6.5 The problem of causal order

On the basis of Axiom II of his first contribution (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 *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 diffeomorphism invariance of 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.

---

<sup>100</sup> Torretti (1978), 391, n.8 rightly observes: 'Riemann's physical geometry requires ideally thin, perfectly flexible and inextensible strings,...

<sup>101</sup> Hilbert (1917), 61.

If  $x_4$  is designated as the ‘proper’ time coordinate, a ‘proper (*eigentlich*) coordinate system’ may be defined as one in which the following 4 inequalities are satisfied by the components of the metric tensor:

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

These ‘reality relations’,<sup>102</sup> implement in the case of general Riemannian geometry the physical requirement of metrical indefiniteness: that 3 of the coordinate axes are space-like, and 1 time-like. 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 space-time description into another proper space-time description is called a *proper* space-time coordinate transformation.

The desired effect of proper coordinate systems is seen in considering a parameterized time-like curve,  $x_s = x_s(p)$ . Since for such a curve, and using Hilbert’s abbreviation,<sup>103</sup>

$$G \left( \frac{dx_s}{dp} \right) < 0, \quad \left[ G \left( \frac{dx_s}{dp} \right) = g_{\mu\nu} \frac{dx_\mu}{dp} \frac{dx_\nu}{dp} \quad (\mu, \nu = 1, 2, 3) \right]$$

it follows that in a proper space-time coordinate system it is always the case that

$$\frac{dx_4}{dp} \neq 0.$$

Accordingly, along such a curve the proper time coordinate continuously increases or decreases. Since any time-like curve remains a time-like 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 space-time coordinate transformation. On the other hand, when  $x_4$  is constant, that is,

$$\frac{dx_4}{dp} = 0,$$

one is dealing with a space-like curve, a *Strecke* or a spatial extension between two world points, for which the first three inequalities of (31) are positive.<sup>104</sup>

<sup>102</sup> Pauli (1921), §22 ‘Realitätsverhältnisse’.

<sup>103</sup> Hilbert (1917), 54.

The significance of these coordinate restrictions for the principle of causality is 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 (31) in addition to our fundamental equations; that is, we restrict ourselves to the use of *proper* space-time coordinates.

As further discussed in section 8, it is not *nature* but the structure of our cognitive experience (in Kantian terms, that of our 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 so-called ‘Gaussian coordinate systems’. Hilbert will employ 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 upon 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.

## 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 contributions: the consequences for causal determination issuing from Theorem 1 of his first contribution. As we have seen, Hilbert regarded this theorem as the ‘guiding motive (*Leitmotiv*) for the construction of my theory’. Theorem 1 stated that, due to the general invariance of the world function stipulated in Hilbert’s Axiom I, there are 4 identities holding 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 December Proofs, but excised entirely in the published version of Hilbert’s

---

<sup>104</sup> Hilbert’s inequalities (31) 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 (31) are merely sufficient, but not necessary, to preserve causal ordering. See Renn and Stachel (1999), 80. Our thanks to a referee for reminding us of this point.

First Note (section 4.2 above), is the *causal underdetermination* of the dynamical evolution of the fields represented by the generally invariant action. We here reiterate that the scope of Theorem 1 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-theoretic concerns of that method, the intent of Theorem 1 was to expose 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 1, in the published version of Hilbert's first paper (and accordingly downplayed in the literature), in his second 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* (31). 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 his second paper is to provide a quite different manner of resolution. Although continuing his interpretation of Theorem 1 that the four generalized Maxwell equations (5) are a consequence of the ten gravitational equations (4), this claim lies well in the background, while the matter of causality takes pride of place. The basic achievement of the paper will be to give the necessary *reformulation* of the causality principle that is required by the new generally invariant field physics.

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. (4) and (5) set up in my first contribution are, as I there especially stressed, in no way of the above-characterized kind. Rather, according to Theorem 1 four are a consequence of the remaining ones: We viewed the four Maxwell equations (5) as a consequence of the ten gravitational equations (4) and therefore have only the 10 essentially independent equations (4) 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.<sup>105</sup>

On the other hand, Hilbert claimed that the situation in the newly emerging 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 (4) and (5), 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 have seen, 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 that of individual statements concerning the spatio-temporal evolution of particular systems or objects. Hilbert therefore revisits 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.

---

<sup>105</sup> Hilbert (1917), 60. 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.

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.*<sup>106</sup>

Renn and Stachel (1999, 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.<sup>107</sup>

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 by 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 aspect 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 upon human understanding (‘physical meaningfulness’), a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense

---

<sup>106</sup> Hilbert (1917), 61.

<sup>107</sup> These points are made explicitly in Hilbert (1916b), probably dated 21 November 1916 (see Sauer (2001), 5), 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, 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.’

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.<sup>108</sup>

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

$$(34) \quad g_{\mu\nu} (\mu, \nu = 1, 2, 3); q_s (s = 1, 2, 3, 4)$$

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 (34) 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 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 4-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  $g_{\mu\nu}$  is bound, according to the fundamental equations (4) and (5) of my first contribution, to the electromagnetic quantities, that is, to matter.<sup>109</sup>

---

<sup>108</sup> Norton (1993), 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.

<sup>109</sup> Hilbert (1917), 63.

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

Everything real (*wirklich*) that transpires in the world is a manifestation of the world-metric. Physical concepts are none other than those of geometry.<sup>110</sup>

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 equations (4) and (5) 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 contribution, Hilbert states that now the physical and geometrical laws, derived 'at one blow' from a variational ('Hamiltonian') principle, show how the geometrical quantities  $g_{\mu\nu}$  -- also the mathematical expression for the phenomena of gravitation -- are connected to the electrodynamic 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),

$$(35) \quad g_{11} = 1, g_{22} = 1, g_{33} = 1, g_{44} = -1; g_{\mu\nu} = 0 (\mu \neq \nu)$$

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

---

<sup>110</sup> Weyl (1918b), 385. For discussion, see Ryckman (2005), chapters 4-6.

Hilbert is first concerned to scrutinize his fundamental equations when the  $g_{\mu\nu}$  adopt the ‘pseudo-Euclidean’ (Minkowski) values (35), and similarly the electrodynamic potentials  $q_s = 0$ . Setting the former into the first term of the fundamental equations (26) – see section 4.1 -- we obtain

$$\left[ \sqrt{g} K \right]_{\mu\nu} = 0,$$

while setting the latter into the second term of the fundamental equations (26) yields

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

This corresponds to the case of a world distant from all electricity (and so 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 by Einstein in the integration of his field equations.<sup>111</sup> Under certain natural assumptions about the perturbative terms, he 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.*<sup>112</sup>

In other words, under these assumptions, the Minkowski metric is the only non-singular metric corresponding to an electricity-free world. Now 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 space-time (curvature) singularity might be given.<sup>113</sup> In point of fact, Hilbert opined that only regular solutions to the fundamental physical equations were immediately represented in reality.<sup>114</sup> As we know, this was also one of Einstein’s most pious beliefs.<sup>115</sup> 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 1916/17, 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, which

---

<sup>111</sup> Einstein (1916b). 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), 279-280.

<sup>112</sup> Hilbert (1917), 66.

<sup>113</sup> Earman (1995) notes that such a characterization (as ‘geodesic incompleteness’) came only in the 1960s in the work of Penrose and Hawking.

<sup>114</sup> Hilbert (1917), 70.

<sup>115</sup> For discussion, see Earman and Eisenstaedt (1999).

Hilbert specified to be one-one and invertible.<sup>116</sup> However, subsequent work showed that Hilbert's definition of a non-singular space-time metric is inadequate, not succeeding in distinguishing true curvature singularities from those arising from coordinate transformation.<sup>117</sup> This must be taken into account in his discussion of the Schwarzschild singularity.

### **6.7.2 Re-derivation 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 GR. This was the first exact solution of the field equations. Still earlier in Nov. 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 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 mass point 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, in 1917, Einstein achieved the same result in explicit cosmological terms with his 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 December Proofs concerns Hilbert's first attempt to resolve the tension between causality and general covariance, precisely formulated by Theorem I. In the 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 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

---

<sup>116</sup> Winter Semester Lectures 1916/1917, 118. See the discussion in Eisenstaedt (1989), 218, and in Renn and Stachel (1999), 87-88.

<sup>117</sup> E.g., Earman (1995), 6.

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.<sup>118</sup> It is, presumably, this convergence that led Hilbert to republish, in 1924, both of his contributions 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 space-time coordinates? That Hilbert’s 4 non-covariant energy equations do implicate such a misunderstanding is affirmed by Renn and Stachel:

...we will discuss Hilbert’s treatment of the problem of causality in Paper 2 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.<sup>119</sup>

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 issue 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 evidence to this effect.

Consider first the direct evidence: Hilbert’s Theorem I as compared with Einstein’s ‘hole argument’. These can be considered different construals of the same problem *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.<sup>120</sup> 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 generally

---

<sup>118</sup> See Hilbert (1923), and Majer and Sauer (2005, 263), who state that Hilbert identified the differences with Einstein ‘as being of a purely nominal nature’.

<sup>119</sup> Renn and Stachel (1999), 77 and 83.

<sup>120</sup> As Brown and Brading (2002), 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.’

covariant gravitational theory was not possible because of failure of causal determination. After all, even with his 4 supplementary energy equations, the field equations of Hilbert's gravitational theory, even if only written 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',<sup>121</sup> 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 4 linear combinations of the Euler-Lagrange brackets are divergences (these are 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<sup>122</sup> laid out an entirely different route than was 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 space-time coordinates.

Second, there is Hilbert's strategy, exactly opposed 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, Hilbert gave many indications that the Gordian knot was to be cut by surrendering causality -- in the sense of Cauchy determination -- not general covariance.<sup>123</sup> 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's Second Communication expressly takes up the causal principle in the physical 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.<sup>124</sup> Once the axiomatic method had mathematically demonstrated a *prima facie* antagonism between two pillars of objectivity, one going back to Kant (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 of 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),

---

<sup>121</sup> Howard and Norton (1993).

<sup>122</sup> And, to a lesser extent, the initial value problem.

<sup>123</sup> See the 'Causality Lecture', (1916b), and the WS (1916/17) Lectures on 'Foundations of Physics'.

<sup>124</sup> See e.g., Hilbert (1919-20), 86-7.

but rather sought his solution in finding what further conditions must 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 space-time coordinates. Hilbert's problem was a different problem, and one that the 'point coincidence' solution to Einstein's problem 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 upon (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 version, 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 has recently remarked that

Hilbert's preference for Kantianism was largely incompatible with Kant's philosophy.<sup>125</sup>

However, this should be taken with a grain of salt, for it largely depends on what is meant by 'Kant's philosophy'. To many of the neo-Kantians active in the first quarter of the 20<sup>th</sup> century, Kant's philosophy was largely a work in progress, not a finished edifice, and nearly all were prepared to concede, as was Hilbert, that 'Kant greatly overestimated the role and the extent of the *a priori*.'<sup>126</sup> By the same token, it might be said that Hilbert, through the axiomatic method, was the first neo-Kantian to put his finger 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.

---

<sup>125</sup> Peckhaus (1994), 91

<sup>126</sup> Hilbert (1930), 383; Engl. transl., 1162.

Hilbert essentially argued that this is no longer the case 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 placed invariance as the 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).<sup>127</sup>

For this the 'axiomatic method' is ideally suited, for it works up a domain of intuitively presented facts into a purely logical-formal system of relations. Space, time, (and causality, which presupposes both) are necessary moments of human cognition, required also as conditions of the possibility of observation and measurement in science. Thus, 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. While the goal of physical science remains that of a completely objective description of nature, a systematically unified, observer-free cognition of nature, this requires an emancipation (never fully attainable) of our view of the processes of nature from all the subjective elements pertaining to human sensibility. The principal step along this new path is 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

---

<sup>127</sup> Hilbert (1919-20), 49.

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.<sup>128</sup>

In thus referring to 'Einstein's theory', it is clear that Hilbert regarded general covariance as its philosophical focal point, as a supreme 'principle of objectivity', and 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*).<sup>129</sup>

Physical science, in accordance with 'the general principle of relativity' (i.e., general covariance) 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'. Given its axiomatic status, 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.<sup>130</sup>

Of course, for Kant, such ideas are the product of reason, not the faculty of understanding, and we recall that the Kantian epigram chosen for the *Grundlagen der Geometrie*, found at the conclusion of the 'Transcendental Dialectic', simply summarizes what, after all, it has been the broad concern of the 'Transcendental Dialectic' to show, that the progression 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)

Now the 'Transcendental Dialectic' adds to the Kantian account of cognition of the 'Transcendental Analytic' (the first part of the *Critique of Pure Reason*, pertaining to sensibility and the categories of the understanding) precisely the cognitive role of 'theoretical reason' in its 'critical' employment. Kant viewed reason in general (both 'theoretical' and 'practical') as rooted in the human capacity to project beyond given

---

<sup>128</sup> Hilbert, 1921, *Grundgedanken der Relativitätstheorie*, lectures in SS 1921, ed. by P. Bernays; cited by Majer, 1995, 284.

<sup>129</sup> Hilbert (1923), Lecture 1, p. 16.

<sup>130</sup> Hilbert (1925), 392.

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 for natural science: reason’s capacity to surpass the confines of experience through the hypothetical adoption of maxims of systematic unity or ‘unity of nature’.<sup>131</sup> 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).

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*’. As Majer in particular has emphasized, in 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.<sup>132</sup> While Hilbert allowed that such elements could be introduced either through mathematical construction or as axioms,<sup>133</sup> 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 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 space-time geometry becomes a dynamical component of the physical theory.

---

<sup>131</sup> 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’.

<sup>132</sup> Majer (1993).

<sup>133</sup> E.g., Hilbert (1919-20), 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.’

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 4 inequalities (30) for the  $g_{\mu\nu}$ , required that events on a given time-like world line always stand in an unambiguous causal relation. So these inequalities for the  $g_{\mu\nu}$ , essentially specifying that  $x_4$  is time-like, 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 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*. Hilbert later termed his revised understanding of the *a priori* 'the finite attitude', 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 attitude 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 other than 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* attitude (*apriorische Einstellung*) is left, which also lies at the foundation of pure mathematical knowledge: it is essentially that finite attitude characterized by me in different essays.<sup>134</sup>

## 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

---

<sup>134</sup> Hilbert, (1930), 383, 385.

covariance appearing as a problem of underdetermination. The material cut from the December Proofs is his first attempt at resolving this tension.

The Second Communication 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 is no longer adequately representable through the senses or by intuition. Hilbert exalts this new requirement on objectivity by according it axiomatic standing.

Causality, 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 upon human understanding, a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense qualities, the 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 Contribution 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 (31) 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 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'.

## Acknowledgements

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. Finally, one of us (K. B.) would like to thank Wolfson College, Oxford, for their support during the early years of the research behind this paper.

## References

- Adler, R., M. Bazin, and M. Schiffer. (1975). *Introduction to General Relativity*. 2<sup>nd</sup> edition. New York: McGraw-Hill.
- Born, M. (1914). “*Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie*”, *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalisch Klasse*, 23-36.
- \_\_\_\_\_. (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 and J. Eisenstaedt (eds.), *The Universe of General Relativity (Einstein Studies v. 11)*. Boston/Basel/Berlin: Birkhäuser, 125-35.
- Brown, H. and K. Brading (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 J. Gray (1999), 145-188
- \_\_\_\_\_. (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., J. Renn, and J. Stachel (1997). “Belated Decision in the Hilbert-Einstein Priority Dispute”, *Science* 278, 1270-73.

Earman, J. (1995). *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. New York: Oxford University Press.

Earman, J. and J. Eisenstaedt (1999). "Einstein and Singularities", *Studies in History and Philosophy of Modern Physics* 30B, 185-235.

Earman, J. and C. Glymour. (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), 245-8.

\_\_\_\_\_. (1916a). "*Die Grundlage der allgemeinen Relativitätstheorie*", *Annalen der Physik* 49, 769-822. Reprinted in Einstein (1996), 284-337. Translated by W. Parrett and G. Jeffrey as "The Foundation of the General Theory of Relativity", in H. Lorentz *et.al* (1923), 109-65.

\_\_\_\_\_. (1916b). "*Näherungsweise Integration der Feldgleichungen der Gravitation*", *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse*, 688-696. Reprinted in Einstein (1996), 348-56.

\_\_\_\_\_. (1917). "*Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*", *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse*, 142-52. Reprinted in Einstein (1996), 541-52. Translated by W. Parrett and G. Jeffrey as "Cosmological Considerations on the General Theory of Relativity", in H. Lorentz *et.al* (1923), 175-88.

\_\_\_\_\_. (1921). "*Geometrie und Erfahrung*", *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte. Physikalisch-mathematischen Klasse*, 123-30. Issued separately in expanded form, Berlin: J. Springer. Reprinted in Einstein (2002), 382-405. Latter translated by W. Parrett and G. Jeffrey as "Geometry and Experience" in *Sidelights on Relativity*. New York: E.P. Dutton, 1923, 235-56. Reprint New York: Dover, 1983, 27-56.

\_\_\_\_\_. (1936). "*Physik und Realität*", *Journal of the Franklin Institute* 221, 313-337. Translated by S. Bargmann as "Physics and Reality" in Einstein, *Ideas and Opinions*, NY, Crown Publishers, 290-323.

\_\_\_\_\_. (1993). *The Collected Papers of Albert Einstein*, v.5. M. Klein, A. Kox, and R. Schulmann (eds.) Princeton: Princeton University Press.

\_\_\_\_\_. (1995). *The Collected Papers of Albert Einstein*, v.4. M. Klein, A. Kox, J. Renn, and R. Schulmann (eds.)

\_\_\_\_\_. (1996). *The Collected Papers of Albert Einstein*, v.6. M. Klein, A. Kox, and R. Schulmann (eds.) Princeton: Princeton University Press.

\_\_\_\_\_. (1998). *The Collected Papers of Albert Einstein*, v.8A. R. Schulmann, A. Kox, M. Janssen, and J. Illy (eds.) Princeton: Princeton University Press.

\_\_\_\_\_. (2002). *The Collected Papers of Albert Einstein*, v.7. M. Janssen, R. Schulmann, J. Illy, C. Lehner, and D. Buchwald (eds.) Princeton: Princeton University Press.

Einstein, A. and M. Grossmann (1913). *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig und Berlin: B.G. Teubner. Reprinted in Einstein (1995), 303-39.

Eisenstaedt, J. (1989). “The Early Interpretation of the Schwarzschild Solution”, in D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity*. (*Einstein Studies*, v. 1) Boston/Basel/Berlin: Birkhäuser, 213-33.

Geroch, R. (1966). “Electromagnetism as an Aspect of Geometry? Already Unified Field Theory – The Null Field Case”, *Annals of Physics* 36, 147-87.

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 and L. Witten (eds.), *Relativity: Proceedings of the Relativity Conference in the Midwest, Cincinnati, Ohio, June 2-6, 1969*. New York: Plenum Press, 161-207.

Hallett, M. (1994). “Hilbert’s Axiomatic Method and the Laws of Thought”, in A. George (ed.), *Mathematics and Mind*. New York: Oxford University Press, 158-200.

Hallett, M. and U. Majer. (2004). *David Hilbert’s Lectures on the Foundations of Geometry 1891-1902*. Berlin/Heidelberg/New York: Springer-Verlag.

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

\_\_\_\_\_. (1900). “*Mathematische Probleme*”. *Vortrag, gehalten auf dem Internationalen Mathematikerkongreß zu Paris 1900*. Reprinted in Hilbert (1935), 290-325. Translation as “Hilbert’s Lecture at the International Congress of Mathematicians” in J. Gray, *The Hilbert Challenge*. Oxford: Oxford University Press, 2000, 240-82.

\_\_\_\_\_. (1915a). “*Die Grundlagen der Physik (Erste Mitteilung)*”. Annotated “*Erste Korrektur meiner erste Note*”, printer’s stamp date “6 Dez. 1915”. 13 pages with omissions.

\_\_\_\_\_. (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 and D. Fine (eds.), *100 Years of Gravity and Accelerated Frames: The Deepest Insights of Einstein and Yang-Mills*. Singapore: World Scientific, 2005, 120-131.

\_\_\_\_\_. (1916a). “*Die Grundlagen der Physik*”. Typescript of summer semester lecture notes. Bibliothek des Mathematisches Institut, Universität Göttingen. 111 pages.

\_\_\_\_\_. (1916b). “*Das Kausalitätsprinzip in der Physik*”. Typescript of lectures, dated 21 and 28 November 1916 by Sauer (2001). Bibliothek des Mathematisches Institut, Universität Göttingen. 17 pages.

\_\_\_\_\_. (1916-17). “*Die Grundlagen der Physik II*”. Typescript of winter semester lecture notes, annotated by R. Bär. Bibliothek des Mathematisches Institut, Universität Göttingen. 199 pages.

\_\_\_\_\_. (1917). “*Die Grundlagen der Physik (Zweite Mitteilung)*”. *Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematische-Phyikalische Klasse*, 53-76.

\_\_\_\_\_. (1918). “*Axiomatisches Denken*”, *Mathematische Annalen* 78, 405-15; reprinted in Hilbert (1935), 146-56. Translation by W. Ewald as “Axiomatic Thought” in W. Ewald, *From Kant to Hilbert: A Sourcebook in the Foundations of Mathematics*, v. 2. New York: Oxford University Press, 1107-15.

\_\_\_\_\_. (1919-20). *Natur und mathmatisches Erkennen. Vorlegungen, gehalten 1919-1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays*. D. Rowe (ed.), Basel-Boston-Berlin: Birkhäuser, 1992.

\_\_\_\_\_. (1922). “*Neubegründung der Mathematik. Erste Mitteilung*”, *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1, 157-77. Reprinted in Hilbert (1935), 157-77. 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*, v. 2. New York: Oxford University Press, 1117-34.

\_\_\_\_\_. (1923). “*Die Weltgleichungen*”. Three lectures given in summer 1923 in Hamburg. To appear in a forthcoming volume published by Springer Verlag on Hilbert’s lectures on physics, edited by U. Majer and T. Sauer.

- \_\_\_\_\_. (1924). “Die Grundlagen der Physik”, *Mathematische Annalen*, 92, 1-32.  
Reprinted with alternations and annotations in Hilbert (1935), 258-89.
- \_\_\_\_\_. (1926). “Über das Unendliche”, *Mathematische Annalen* 95, 161-90.  
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, 367-392.
- \_\_\_\_\_. (1930). “Naturerkennen und Logik”, *Die Naturwissenschaften* 17, 959-963;  
reprinted in Hilbert (1935), 378-387.
- \_\_\_\_\_. (1935). *Gesammelte Abhandlungen*. Bd. III. Berlin: J. Springer.
- Hofer, C. (2000). “Energy Conservation in GTR”, *Studies in History and Philosophy of Modern Physics* 31, 187-199.
- Howard, D. and J. Norton (1993). “Out of the Labyrinth? Einstein, Hertz, and the Göttingen Answer to the Hole Argument”, in J. Earman, M. Janssen, and J. Norton (eds.), *The Attraction of Gravitation: New Studies in the History of General Relativity (Einstein Studies v. 5)*. Boston/Basel/Berlin: Birkhäuser, 30-62.
- Janssen, M. and J. Renn (2007). “Untying the Knot: How Einstein Found his Way Back to Field Equations Discarded in the Zürich Notebook”, in J. Renn *et al.* (eds.), *The Genesis of General Relativity: Documents and Interpretations*. Dordrecht: Springer, pages.
- Klein, F. (1917). “Zu Hilberts erster Note über die Grundlagen der Physik”, *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse*. 469-82. Reprinted with additional remarks in Klein, *Gesammelte Abhandlungen*, Bd.I. Berlin: J. Springer, 1921, 553-567.
- \_\_\_\_\_. (1927). *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*. Teil 2. Berlin: J. Springer.
- Landau, L. and E. Lifshitz. (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. (1993). “Hilberts Methode der Idealen Elemente und Kants regulativer Gebrauch der Ideen”, *Kant-Studien* 84, 51-77.
- \_\_\_\_\_. (1995). “Geometry, Intuition and Experience: From Kant to Husserl”, *Erkenntnis* 42, 261-85.

\_\_\_\_\_. (2001). "The Axiomatic Method and the Foundations of Science: Historical Roots of Mathematical Physics in Göttingen (1900-1930)", in M. Rédei and M. Stöltzner (eds.), *John von Neumann and the Foundations of Quantum Physics*. Dordrecht: Kluwer, 11-33.

Majer, U. and T. Sauer. (2004). "Hilbert's 'World Equations' and his Vision of a Unified Science", in A. J. Kox and J. Eisenstaedt (eds.), *The Universe of General Relativity (Einstein Studies v. 11)*. Boston/Basel/Berlin: Birkhäuser, 259-76.

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, 511-34; "Zweite Mitteilung", *Annalen der Physik* 39, 1-40.

\_\_\_\_\_. (1913). "Grundlagen einer Theorie der Materie. Dritte Mitteilung", *Annalen der Physik* 40, 1-66.

\_\_\_\_\_. (1917). "Die Einsteinsche Gravitationstheorie und das Problem der Materie", *Physikalische Zeitschrift* 18, 551-6, 574-80, 596-602.

Misner, C. and J. Wheeler (1957). "Classical Physics as Geometry", in J. Wheeler, *Geometrodynamics*. New York: Academic Press, 1962, 225-307.

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, 89-124.

Noether, E. (1918). "Invariante Variationsprobleme", *Nachrichten. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematische-Physikalische Klasse*, 235-57. Translation by M. Tavel as "Noether's Theorem", in *Transport Theory and Statistical Physics* 1 (1971), 183-207.

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

\_\_\_\_\_. (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 jr., W. (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.

\_\_\_\_\_. (1979). *Wolfgang Pauli. Wissenschaftlicher Briefwechsel*. Bd. I. A. Hermann, K. v. Meyenn, and V. Weisskopf (eds.), Heidelberg/Berlin/New York: Springer Verlag.

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

\_\_\_\_\_. (1994). “Hilbert’s Axiomatic Programme and Philosophy”, in E. Knoblauch and D. Rowe (eds.), *The History of Modern Mathematics*, v.3. New York: Academic Press, 91-112.

Pyenson, L. (1985). “Relativity in Late Wilhelmian Germany: The Appeal to a Pre-established Harmony between Mathematics and Physics”, in Pyenson, *The Young Einstein*. Bristol: Adam Hilger Ltd., 137-155.

Rainich, G. (1925). “Electricity in Curved Space-Time”, *Nature* 115 (1925), 498.

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

Renn, J. and M. Schemmel (eds.) 2007. *The Genesis of General Relativity. Vol. 4. Gravitation in the Twilight of Classical Physics: The Promise of Mathematics*. Springer.

Renn, J. and J. Stachel (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. Now available in Renn and Schemmel (eds.), 2007, 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*. New York: Oxford University Press, 189-233.

\_\_\_\_\_. (2001). “Einstein Meets Hilbert: At the Crossroads of Mathematics and Physics”, *Physics in Perspective* 3, 379-424.

\_\_\_\_\_. (2002). “Einstein’s Gravitational Field Equations and the Bianchi Identities”, *The Mathematical Intelligencer* 24, 57-66.

\_\_\_\_\_. (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, 529-75.

\_\_\_\_\_. (2001). "The Relativity of Elaboration: Hilbert's Second Note on the Foundations of Physics", ms. dated 10 September 2001. 14 pp.

\_\_\_\_\_. (2002). "Hopes and Disappointments in Hilbert's Axiomatic 'Foundations of Physics'", in M. Heidelberger and F. Stadler (eds.), *History of Philosophy of Science: New Trends and Perspectives*. Dordrecht: Kluwer, 225-37.

\_\_\_\_\_. (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*

Stachel, J. (1992). "The Cauchy Problem in General Relativity – The Early Years", in J. Eisenstaedt and A.J. Kox (eds.), *Studies in the History of General Relativity*. (Einstein Studies, v. 3) Boston/Basel/Berlin: Birkhäuser, 407-418.

\_\_\_\_\_. (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*. University of Pittsburgh Press, Pittsburgh, 129-160.

\_\_\_\_\_. (1999). "New Light on the Einstein-Hilbert Priority Question", *Journal of Astrophysics and Astronomy* 20, 91-101; as reprinted in Stachel (2002), 353-364.

\_\_\_\_\_. (2002). *Einstein from 'B' to 'Z'*. Boston/Basel/Berlin: Birkhäuser. (*Einstein Studies*, v.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*. New York: John Wiley, 169-98.

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*, 334-344.

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

Vizgin, V. (1994). *Unified Field Theories in the First Third of the 20<sup>th</sup> Century*. Translated from the Russian by J. Barbour. Boston/Basel/Berlin: Birkhäuser.

\_\_\_\_\_. (2001). "On the Discovery of the Gravitational Field Equations by Einstein and Hilbert: New Materials", *Physics-Uspekhi* 44 (12), 1283-1298.

Weyl, H. (1918a). *Raum-Zeit-Materie*. Berlin: J. Springer.

\_\_\_\_\_. (1918b). "Reine Infinitesimalgeometrie", *Mathematische Zeitschrift* 2, 384-411. Reprinted in Weyl, *Gesammelte Abhandlungen* Bd. 2. Berlin: Springer Verlag, 1968, 1-28.

\_\_\_\_\_. (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* 28, 147-240.

## Appendix 1: A sketch of the Cauchy problem in general relativity

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

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

$$(1) \quad g_{mn}, \quad g_{mn,4} \quad (m, n = 1, 2, 3)$$

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

$$(2) \quad g_{mn, \mu} \quad g_{mn, \mu\nu} \quad g_{mn, 4\mu}.$$

Expressions (1) and (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

$$(3a) \quad R_{mn} = \frac{1}{2} g^{44} g_{mn,44} + M_{mn} = 0$$

$$(3b) \quad R_{m4} = -\frac{1}{2} g^{4n} g_{mn,44} + M_{m4} = 0$$

$$(3c) \quad R_{44} = \frac{1}{2} g^{mn} g_{mn,44} + M_{44} = 0,$$

where  $m, n = 1, 2, 3$ , and the  $M_{\mu\nu}$  can be expressed in terms of the initial data on  $\Sigma$  (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  $g_{\mu 4, 44}$  ( $\mu = 1 - 4$ ) needed to describe the time evolution of the metric from the data on  $\Sigma$ . 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  $\Sigma$  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 space-time geometry. It is to remedy this underdetermination that Hilbert will prescribe 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

$$(4) \quad g_{\mu 4} = 0, \quad g_{44} = 1,$$

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

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

$$(5a) \quad R_{mn} = 0, \quad \text{and}$$

$$(5b) \quad G_{\mu}{}^4 = G^4{}_{\mu} = 0.$$

Then the initial data cannot be given arbitrarily on  $\Sigma$  but must satisfy the four 'constraint equations' (5b). These four gravitational field equations are independent of the  $g_{\mu\nu,44}$  and must be solved for the hypersurface  $\Sigma$  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 space-like surface  $\Sigma$  ( $x^4 = x'^4 = \text{const}$ ), 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  $G_{\mu}{}^4 = 0$ . 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  $R_{\mu\nu} = 0$ .

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

$$g'_{mn} = g_{mn} + g_{mn,4} dx'^4, \quad \text{and}$$

$$g'_{mn,4} = g_{mn,4} + g_{mn,44} dx'^4.$$

These quantities are the initial data on  $\Sigma'$ . The constraint equations (5b) are automatically satisfied on  $\Sigma'$  since, if they are satisfied on  $\Sigma$ , then  $G^n_m = 0$ , and  $G^4_{m,4} = 0$ , so

$$G'^4_m = G^4_m + G^4_{m,4} dx'^4 = 0.$$

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

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

## Appendix 2: 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