8

Hilbert's Axiomatic Method and His "Foundations of Physics": Reconciling Causality with the Axiom of General Invariance

Katherine A. Brading and Thomas A. Ryckman

University of Notre Dame and Stanford University, USA

So it is that all human cognition begins with intuitions,
proceeds from there to concepts, and ends with ideas.
Kant, *Critique of Pure Reason.* (A702/B730)
Epigram to Hilbert (1899)

8.1 Introduction

In November and December 1915, Hilbert gave two presentations to the Royal Göttingen Academy of Sciences under the common title 'The Foundations of Physics'. Distinguished as 'First Communication' (Hilbert, 1915b) and 'Second Communication' (Hilbert, 1917), the two 'notes', as they are widely known, eventually appeared in the *Nachrichten* of the Academy. The first quickly entered the canon of classical general relativity but has recently become the object of renewed scholarly scrutiny since the discovery (Corry, Renn and Stachel 1997) of a set of printer's proofs dated December 6, 1915 ((Hilbert, 1915a), henceforth 'Proofs'). Hilbert's second presentation has not received the same detailed reconsideration, with the recent exception of an extended study offered by Renn and Stachel (1999/2007). While we agree with much of their detailed technical reconstruction, we profoundly disagree with the assessment of Renn and Stachel that the second note shows that Hilbert had abandoned his own project (set out in the first note), and is working on a variety of largely unrelated problems within Einstein's. In our opinion, this assessment rests on misunderstandings concerning the aims, content, and significance of the second communication, as well as its links to the first. Our aim in this paper is to offer an alternate narrative, according to which Hilbert's second note emerges as a natural continuation of the first, containing important and interesting further developments of that project, and above all shedding needed illumination on Hilbert's assessment of the epistemological novelty posed by a generally covariant physics.

Hilbert's notes on 'Foundations of Physics' traditionally have been assessed solely in terms of the contributions they made to general relativity, as that theory is

C. Lehner et al. (eds.), *Einstein and the Changing Worldviews of Physics*, Einstein Studies 12, DOI 10.1007/978-0-8176-4940-1_8, © 2012 The Center for Einstein Studies

known in its completed form. 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 invariance in one blinding flash, masterfully wielding an arsenal of advanced mathematics. Our main contention is that such assessments radically occlude internal motivations, which are largely logical and epistemological, and so cast them in a misleading light. In particular, the explicitly stated epistemological intent of the 'axiomatic method' is ignored, as are Hilbert's own express assertions regarding his construction as a triumph of that method. But set within the 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. Unlike Einstein's largely informal and heuristic extraction from the clutches of the hole argument, Hilbert stated the difficulty in a mathematically precise manner as an ill-posed initial value problem, 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 note, was the 'point of departure' for his axiomatic investigation.

8.2 The Essential Context: Hilbert's Axiomatic Method and Kantian Epistemology

Hilbert's first note opens with a declaration that the ensuing 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 obtained redound 'certainly to the most magnificent glory of the axiomatic method.' Unless mere rhetorical embellishment, these passages establish that the 'axiomatic method' (whatever that may be) played an integral role in the enterprise at hand. Understanding the significance of Hilbert's setting his results squarely within the frame of the axiomatic method is accordingly essential.

What, then, is the axiomatic method? In the literature, it has been widely assumed that Hilbert's references to 'axiomatic method' simply signal his derivation of 14 fundamental field equations, as well as several subsidiary theorems, from two principal axioms (e.g., Guth, 1970, 84; Mehra, 1974, 26, 72 n. 145; Pais, 1983, 257). However, as can be documented in numerous lecture courses going back at least to 1905, the term not only implicates a typical mathematical concern with the rigorous explicit statement of a theory, but also connotes a specifically *logical and epistemological* method of investigation of mathematical theories (including those of physics) that Hilbert pioneered, and which he saw as closely tied to the nature of thought itself.[1]

In published articulation, the axiomatic method debuted in Hilbert's classic Gauss–Weber *Festschrift* essay, *Grundlagen der Geometrie* (1899). The essay's epigraph has been little noticed, yet is worth quoting in the original German, for it is Kant's most concise statement of how cognition requires, and results 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)

As Kant's directive prescribes, the axiomatic method is conceived as a logical analysis of cognitions 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'.[2] 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. The axioms standing at the pinnacle of the *Fachwerk von Begriffen* are not only general but also *ideal*. They are, as far as possible, independent of the particular intuitions (and so, concrete facts) from which the process started. By virtue of their ideality, and thus their severance from experience and intuition, 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, is thereby highlighted. Axioms thus play a hypothetical or guiding role in cognition. As will be seen, Hilbert considered axioms to be 'things of thought' or indeed, 'ideas' in Kant's regulative sense, effecting a separation between logical/mathematical and intuitional/experiential thought, even as the latter has thus been arranged in deductive form. Indeed, it is just 'the service of axiomatics'

> to have stressed a separation into the things of thought *(die gedanklichen Dinge)* of the (axiomatic) framework and the real things of the actual world, and then to have carried this through.[3]

Use of the axiomatic method does not aim, at least in the first instance, at the discovery or recognition of *new* laws or principles, but at the conceptual and logical

[1] 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.'

[2] Hilbert (1922); 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 a *spatial* or *temporal* intuition.

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

clarification or reconstruction of known ones.[4] Ultimately, the axiomatic method is concerned with demonstrating that the axioms selected for a theory possess the three meta-logical properties or relations of mutual consistency, independence, and completeness.[5] Combining 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.[6]

8.2.1 Mie's Theory and the Axiomatic Method

We recall that the task of the axiomatization of physical theories was the sixth in the famous list of 23 mathematical problems Hilbert posed at the 1900 International Congress of Mathematicians in Paris. Inclusion of the axiomatization of physics among the other purely mathematical problems appears rather incongruous until Hilbert's lifelong interest in physics is taken into account.[7]

> 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 must, with a small number of axioms, try to include as large a class of physical phenomena as possible, 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 exactly in each instance whether the new axioms are compatible with the previous ones. The physicist, as his theories develop, often finds himself forced by the results of his experiments to make new hypotheses, while he depends, with respect to the compatibility of the new hypotheses with the old axioms, solely upon these experiments or upon a certain physical intuition, a practice which is not admissible in the rigorously logical construction of

[4] See Majer (2001), 19.

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

[6] Hilbert (1918), 407 (Engl. trans., 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.'

[7] 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.

> a theory. The desired proof of the compatibility of all assumptions seems to me also of importance, because the effort to obtain such proof always forces us most effectively toward an exact formulation of the axioms.[8]

Three items of interest mark this passage.

- 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 theories 'near to reality', and so the axiomatic method is ideally suited for setting up a speculative or hypothetical theory.
- 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 construction of a theory' in place of its informal statement in experiential or intuitive terms.

Above all, we wish to stress the *hypothetical* character of Hilbert's axiomatic approach to physics. This aspect was explicitly underlined by Hilbert's former student and Göttingen physics colleague, Max Born, in a tribute on the occasion of Hilbert's 60th birthday.

> [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.[9]

These points are of special significance for understanding the role of the Mie theory in Hilbert's two notes on the 'Foundations of Physics'.

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. Hilbert knew of the Mie theory at least since the discussion of it at the Göttingen Mathematical Society in December 1912, and again in December 1913, when Born had put it into a more canonical mathematical form (Corry, 1999, 176). Certainly, the fact that Mie had sought to derive field equations of a generalized Maxwellian electrodynamics from the axiom of a Lorentz invariant 'world function' fitted very naturally into Hilbert's axiomatic approach. A central attraction of the Mie theory was that then, coupled with Einstein's theory of gravitation, it enabled a *hypothetical axiomatic completion*

---

[8] As translated in Gray (2000), 257–258.

[9] Born (1922), 90–91, our translation. Unless otherwise noted, all translations in this paper are our own.

of physics that could be studied by drawing consequences from the amalgamation of the two theories. In this 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 framework of concepts so that it happens that the objects of the field of knowledge correspond to the concepts, and the assertions regarding the objects to the logical relations between the concepts. Through this mapping, the (logical) investigation becomes entirely detached from concrete reality *(Wirklickkeit)*. The theory has nothing more to do with real objects *(realen Objekten)* or with the intuitive content of knowledge. It becomes a pure construction of thought *(reine Gedankengebilde)*, of which one can no longer say that it is true or false. Nevertheless, this framework of concepts has 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.[10]

Yet the Mie theory was attractive for a number of other mathematical and philosophical reasons that merit illumination. 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 the summer of 1913 and would again in the summer of 1917 (Corry, 1999, 174, 183). 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 (Hilbert, 1916a, 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 electrodynamical equations for the four unknown electrodynamic potentials. From given boundary and initial conditions, one could show that the state of the world at any future time could be univocally determined via these equations through specification of the values of these potentials at any prior time, as required by the principle of causality (as Hilbert understood that principle). Notoriously, the Mie theory purchases its causal determination at the cost of gauge invariance (the Mie potentials have 'absolute' values). Ironically, what current wisdom deems precisely wrong about the Mie theory was thus a philosophical ground in favor of it cited by Hilbert.[11] In sum, in the summer of 1916, and so

[10] Hilbert's WS 1921/1922 Lectures on the '*Grundlagen der Mathematik*,' as cited and translated in Hallett (1994), 167–168.

[11] 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.

after Einstein's canonical presentation of general relativity (Einstein, 1916), Hilbert continued to regard the standing of the principle of causality in the new physics of Einstein's principle of general invariance as unclear (Hilbert, 1916a, 110). Mie's theory, however, was deemed suitable to be incorporated into Hilbert's axiomatic construction by its *a priori* consistency with the requirement of causality. Finally, we shall see that there were also *a posteriori* reasons justifying Hilbert's incorporation of Mie's theory. Namely, Hilbert would 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. Both of these results are crucial to Hilbert's otherwise problematic claim that electrodynamic phenomena are a consequence of gravitation.

8.3 Hilbert's First Note: What Was Hilbert's Aim?

As legend has it, in November 1915, Hilbert engaged with Einstein in a competition 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, for he postulated an action integral containing a Lagrangian 'world function' 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 notes (1915b, 395; 1917, 63–64).

The first note accordingly begins with recognition that the investigations of Einstein and Mie have 'opened new paths for the investigation of the foundation of physics'. 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 four electromagnetic vector potentials q_s, Hilbert employed sophisticated mathematical techniques to draw out the consequences of his two principal axioms. It is clear that Hilbert was extremely pleased with the axiomatic conjunction of the two theories. The triumphal language at the end of his first note expresses 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.

8.3.1 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' (Lagrangian density) H, depending upon the 10 gravitational potentials $g_{\mu\nu}$, their first and second derivatives, as well as the four electromagnetic potentials q_s, and their first derivatives:

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

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

The function H is not further specified. But 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 note, and as Mie noted that same year,[12] 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 his first note demonstrated. In both cases, objective significance will accrue only to objects invariant under arbitrary transformation of the parameters, respectively, coordinates. Precisely the same language of 'world parameters' is also used in the Proofs, *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 his second 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 (1915a, 2; 1915b, 396).

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 (Hilbert, 1924, 4).

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

[12] 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.

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

- THEOREM I ('*Leitmotiv*'). In the system of n Euler–Lagrange differential equations in n variables obtained from a generally covariant variational integral such as in Axiom I, 4 of the n equations are always a consequence of the other $n-4$ in the sense that 4 linearly independent combinations of the n equations and their total derivatives are always identically satisfied.[13]

One of Hilbert's principal claims, to be discussed below, is that, as a consequence of Theorem I, electromagnetic phenomena may be regarded as consequences of the gravitational. The theorem also gives rise to Hilbert's 'problem of causality' (see Section 8.4.1).

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 ten equations for the gravitational potentials,

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{g_{kl}^{\mu\nu}} = 0, \quad \text{or} \quad [\sqrt{g}\,H]_{\mu\nu} = 0 \tag{8.1}$$

$$\left[g_l^{\mu\nu} = \frac{\partial g^{\mu\nu}}{\partial w_l}; g_{l\,k}^{\mu\nu} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l \partial w_k}\right],$$

while differentiation of H with respect to the electromagnetic potentials q_s and their first derivatives yields four equations,

$$\frac{\partial\sqrt{g}H}{\partial q_h} - \sum_\sigma \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}\,H}{\partial\, q_{h\,k}} = 0, \quad \text{or,}$$

$$[\sqrt{g}H]_h = 0\left[q_{hk} = \frac{\partial q_h}{\partial w_k}\quad (h,\, k = 1,2,3,4)\right].^{14} \tag{8.2}$$

The fourteen equations [(8.1)] and [(8.2)] in Hilbert (1915b) are termed 'the basic equations of gravitation and electrodynamics or generalized Maxwell equations'. On the assumption that the Mie theory renders a viable theory of matter, these equations encompass the entirety of fundamental physics. The remainder of the paper concerns Hilbert's treatment of energy, which includes his demonstration of a connection between the phenomena of gravitation and of electromagnetism. We turn to this issue now.

[13] Hilbert (1915a), 2–3; (1915b), 397. However, as Klein (1917), 481, first pointed out, since Hilbert regards the invariant H as the additive sum of *two* general invariants $K + L$, there are then 8 identities between the 14 field equations. According to Klein, 4 of these are a purely mathematical consequence of the 10 gravitational equations. The other 4 permit Hilbert's interpretation of the electromagnetic equations as a consequence of the gravitational equations.

[14] The form of equations [(8.1)] and [(8.2)] is trivially different algebraically between the Proofs and the published version. Here we follow the published version.

8.3.2 The Connection Between Gravitation and Electromagnetism

On the basis of Theorem I, Hilbert concluded that the four equations [(8.2)] are a consequence of the ten equations [(8.1)], such that, '*in the sense indicated, electrodynamic phenomena are effects of gravitation*' (1915a, 3; 1915b, 397). This claim is certainly not part of the standard lore of general relativity, and it has repeatedly come under severe criticism, most recently by Renn and Stachel (1999, 36–41; 2007, 893–899) and by Corry (2004, 336–337). 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.[15] For present purposes, we wish to highlight three results that Hilbert proudly attributed to the use of the axiomatic method:

- general covariance, as we shall prefer to say, 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.[16] However, we must point out that Hilbert also regarded the second result, concerning the Mie tensor, 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 (1915a, 11; 1915b, 404) $[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0$, where the first term is evaluated, in the published version but not in the Proofs, so that the crucial trace term appears, $[\sqrt{g}K]_{\mu\nu} = \sqrt{g}(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu})$. Now L is a general invariant that, by Axiom I, is assumed to depend *only* on the $g_{\mu\nu}$, the q_s, and their first derivatives $\frac{\partial q_s}{\partial w^l}$. Hilbert had previously shown that, from Axiom II (the axiom of general invariance) and a supporting theorem (Theorem II, the Lie derivative of the metric), it follows that L must satisfy the relations (1915a, 11; 1915b, 403)

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

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

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

[16] Hilbert's considerations on the tension between general covariance and causality are the other central outcome.

$$M_{k\,s} = Rot(q_s) \equiv q_{s\,k} - q_{k\,s},$$

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

> *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.*[17]

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 $\frac{\partial q_s}{\partial w^l}$ enter into L 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.,

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

In this respect, Hilbert's gravitational field equations differ from Einstein's in their interpretation because Hilbert assumed a particular hypothesis about the electromagnetic constitution of all matter.[18] With 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.[19] This is also a fundamental result, and Hilbert emphasized its significance in *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.[20]

Hilbert regarded this result as a central achievement of his theory, and indeed a bellwether of its general correctness. Moreover, it must be emphasized that the '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.

[17] Hilbert (1915b), 403; at the corresponding place in Hilbert (1915a), 10, the bracketed expression does not appear.

[18] Earman and Glymour (1978), 303; Sauer (1999), 564.

[19] For details, see Sauer (1999), 555.

[20] Hilbert (1915a), 10; (1915b), 404. Notice that *already in the Proofs*, Hilbert's reference is to 'Einstein's general theory of relativity' *(der Einsteinschen allgemeinen Relativitätstheorie)*, explicitly according due credit to Einstein.

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 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 (Folgeerscheinuung) 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.*[21]

8.4 Differences Between the Proofs and the Published Version

There are three main differences between the Proofs and the published version of the first note. First, the explicit form of the field equations of general relativity does not appear in the Proofs. This matter has received considerable attention in the recent literature, and will not be treated here. Our view, following the careful analyses of Sauer (1999, 2005), is that the Proofs already contain the correct gravitational field equations of general relativity in the inexplicit form of a variational principle and the Hilbert action. The two other differences are related, and will be treated here: in the Proofs, but entirely missing from the published version, there is a clear statement of the problem of causality facing any generally covariant theory, as well as a proposed solution that restricts the applicability of space-time coordinates. Our interest lies in identifying a thematic link between this text cut from the Proofs and issues treated in the published second note. This enables us to see that it treats in detail the problem of causality that is addressed in the Proofs, but dealt with unsatisfactorily there.

8.4.1 Hilbert's Target: The 'Problem of Causality'

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

[21] 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. 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.

> Our mathematical theorem teaches that for the 14 potentials, the above axioms I and II can yield 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 [(8.1)] and [(8.2)] is essential. (1915a 3–4)

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.

Befitting 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 a well-posed Cauchy problem. Theorem I suggests that it is a property of any such theory that it does not. Prior to general relativity, as Hilbert repeatedly emphasized, all physical theories permitting a variational formulation satisfied Cauchy determination in the sense that they yielded precisely as many independent Euler–Lagrange equations as there were 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, four 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 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 previous physical theories in variational formulation, and so its seeming failure in the context of general invariance surely sparked Hilbert's interest. 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 conditions additional to the ten independent equations that will render the Cauchy problem well posed. Finding the 'missing four space-time' equations is the motivation behind the intricate mathematical construction in the Proofs of an 'energy form'

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

(where e_s is termed the 'energy vector', and p^s is an arbitrary contravariant vector) 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. While by construction this 'energy form' is a general invariant, Hilbert finds four supplementary conditions by imposing special coordinate restrictions on his 'energy vector'; namely, that it must satisfy the divergence equation (numbered (15) in Hilbert, 1915b),

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0, \quad \text{iff the four quantities } \; e_s = 0, \tag{8.3}$$

where the w_l are now special space-time coordinates adapted to this 'energy theorem', as stated in a third, and final axiom appearing only in the Proofs:

- AXIOM III: ('The Axiom of Space and Time'): 'The space-time coordinates are such particular world parameters for which the energy theorem [(8.3)] is valid.'

Elucidating this result, Hilbert clarified the main point, that these four 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. (1915a, 7)

The idea that satisfaction of energy conservation requires four non-covariant conditions is almost certainly taken from Einstein's *Entwurf* theory of 1913–1914 (Renn and Stachel, 1999, 32; 2007, 888). Drawing analogies to Einstein's difficulties in the 'hole argument', Renn and Stachel (1999, 73; 2007, 934) regard Hilbert's energy construction as his 'Proofs argument, based on causality, against general covariance'. But Hilbert's rather more complicated construction has, philosophically and motivationally, a different *raison d'être.*[22] We see that in Hilbert's case, *the aim was to extract a Cauchy-determinate structure within an otherwise generally covariant theory, and not to abandon general covariance.*[23] However, the complex mathematical derivation in the Proofs leading to Hilbert's four energy conditions was

[22] Brading and Ryckman (2008), section 7, emphasize the *disanalogies* of Hilbert's and Einstein's respective treatments of the tension between causality and general covariance.

[23] This is also noted 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.

cut, together with *all* of its motivation, from the published version. In the light of the completed form of Einstein's theory, placing coordinate restrictions on the energy term turned out to be the wrong approach for solving the tension between general covariance and Cauchy determination. Hilbert accordingly dropped it altogether, modifying and truncating his treatment of energy. There—almost certainly following the implicitly generally covariant energy in Einstein's final November presentation to the Berlin Academy (Einstein, 1915)—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 does not go away for Hilbert. We claim that the second note contains his much revised, detailed reconsideration of this issue, and is rightly understood only in this light.[24]

8.5 Hilbert's Second Note

This paper is, we argue, principally concerned with providing a satisfactory reconciliation between the principles of general invariance and causality. The impression given by Renn and Stachel is that the second note is a list of special topics within general relativity. Moreover, they allege that the second note shows Hilbert in 'agony' over the 'collapse of his own research program' (Renn and Stachel, 1999, 90; 2007, 953). On the contrary, in our opinion, Hilbert deemed his second note to be its completion.

The treatment of the problem of causality in generally covariant theories here has four principal facets. First, Hilbert observed that arbitrary point transformations (diffeomorphisms) do not respect the relation of cause and effect among world points lying on the same timelike curve. To rectify this, he introduced the notion of *proper coordinate systems*, transformations among which always respect the distinction between spacelike and timelike coordinate axes and can never reverse the temporal order of cause and effect. Next, he pointed to the consequent need to reformulate the 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, though we shall not discuss it here,[25]

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

[25] See Brading and Ryckman (2008), section 6.

Hilbert also took up the related issue of the inconsistency of Euclidean geometry (permitting, on account of its globally fixed metrical structure, the concept of action at a distance) with the new physics of fields, which he calls a four-dimensional pseudo-geometry. To this end, he discussed the conditions under which a pseudo-Euclidean (Minkowski) metric arises in the new physics, and he rederived the external Schwarzschild solution corresponding to the solar gravitational field without the assumption that the $g_{\mu v}$ had pseudo-Euclidean values at infinity, that is, that the solar system is embedded in a pseudo-Euclidean world.

8.5.1 The Problem of Causal Order

On the basis of Axiom II of his first note, and with implicit reference to Einstein's requirement of general covariance for the gravitational field equations, 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.

Although Hilbert speaks (1917, 57) of 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 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 transformation this causal ordering is preserved. To achieve this end, Hilbert introduced what he called 'proper' coordinate systems.

If x_4 is designated as the 'proper' time coordinate, a 'proper *(eigentlich)* coordinate system' may be defined as one in which the following four inequalities are satisfied by the components of the metric tensor (numbered (31) in Hilbert, 1915b):

$$g_{11} > 0, \quad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0, \quad g_{44} < 0. \tag{8.4}$$

These so-called 'reality relations'[26] implement, in the case of general Riemannian geometry, the physical requirement of metrical indefiniteness: that three of the coordinate axes are spacelike, and one timelike. Together, the restrictions imply that $g(= \det |g_{\mu v}|) < 0$, so $\sqrt{-g}$ must replace $\sqrt{g}$ in all tensor formulae. A coordinate

[26] Pauli (1921), S22 '*Wirklichkeitsverhältnişe*'.

transformation carrying such a proper space-time description into another proper space-time description is called a *proper* space-time coordinate transformation.[27]

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

> So we see that the concepts of cause and effect lying at the basis of the principle of causality also in the new physics never leads to inner contradiction, as soon as we always take the inequalities [(8.4)] in addition to our fundamental equations; that is, we restrict ourselves to the use of *proper* space-time coordinates.

We observe here (a point we shall return to) that it is not *nature* but the structure of our cognitive experience (preservation of causal relations) that leads to the restriction to proper coordinate systems. Coordinate conditions govern possible objects of experience (for Hilbert, these are physical facts as determined by 'measure threads' and 'light clocks') represented in causal relation within spatio-temporal empirical intuition, but *not* possible objects of physics, governed by the ideal requirement of general invariance, according to Axiom II.

8.5.2 The Problem of Univocal Determination

We reiterate that the scope of Theorem I extends to any generally invariant four-dimensional theory, and is thus broader than either general relativity or Hilbert's theory itself. While the conflict between general invariance and causal determination is only implied, via Theorem I, in the published version of Hilbert's first note (and accordingly downplayed in the literature), in his second note, Hilbert nonetheless claimed that there he had 'especially stressed' this fact. One might conjecture that Hilbert had merely forgotten that 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 had not worked. Then, when revising the Proofs in the light of Einstein's published paper of November 25, 1915, it seems he had not yet seen that a solution lay not in four additional non-invariant *equations*, but rather in the coordinate conditions yielding the four *inequalities* [(8.4)]. Uncertain about how the issue was to be resolved, Hilbert had simply buried the entire issue in the published version.

The main point of his second paper is to provide a quite different manner of resolution. Although continuing his interpretation of Theorem I that the four

[27] Hilbert's inequalities [(8.4)] 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 [(8.4)] are merely sufficient, but not necessary, to preserve causal ordering. See Renn and Stachel (1999), 80; (2007), 941–942.

generalized Maxwell equations [(8.2)] are a consequence of the ten gravitational equations [(8.1)], 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 observes that up until the present time all physical theories whose laws are written as differential equations have satisfied the requirement of causality, in the sense of 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 different once the requirement of general invariance is imposed.

> Now the fundamental eqs. [(8.1)] and [(8.2)] set up in my first contribution are, as I there especially stressed, in no way of the above-characterized kind. Rather, according to Theorem I four are a consequence of the remaining ones: We viewed the four Maxwell equations [(8.2)] as a consequence of the ten gravitational equations [(8.1)] and therefore have only the 10 essentially independent equations [(8.1)] for the 14 potentials $g_{\mu v}$ and q_s.
> As soon as we raise the requirement of general invariance for the fundamental equations of physics, the just mentioned circumstance is essential and even necessary.[28]

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.

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 set as 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 only in the context of laws, but also 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

[28] Hilbert (1917), 60. 'Essential and necessary' because the introduction of a Gaussian (and so, *proper*) space-time coordinate system for the 10 potentials, $g_{\mu\nu}(\mu, \nu = 1, 2, 3); q_s(s = 1, 2, 3, 4)$, would result in overdetermination of the system (more than ten independent equations), and thus inconsistency.

when such statements are predictions, i.e., concern the future, Hilbert stipulated that their meanings are to be understood in such a way that the requirement of physical causality (viz., that causes precede their effects) is satisfied.

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

Renn and Stachel (1999, 81; 2007, 942) observe that this is obviously *not* a claim that physically meaningful statements are independent of the choice of a coordinate system. But is this passage evidence for what they go on to suggest, that Hilbert still attaches 'some residual physical meaning to the choice of coordinates'? In our opinion, it is apparent from Hilbert's formulation that 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.[29]

With this new 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 the new point of view, the physical principle of causality, as ensured by the coordinate conditions of a well-posed Cauchy problem, is a lingering but

[29] These points are made explicitly in Hilbert's 'Causality Lecture' (1916b), 5–6, given the probable date of November 21, 1916, in Sauer (2001): '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: '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.'

not eliminable constraint upon human understanding (‘physical meaningfulness’), a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense qualities, Hilbert viewed the requirement of physical causality as anthropomorphic, having to do not with the objective world of physics but rather with our experience of that world.[30]

8.6 Hilbert’s Revision of Kant in the Light of General Invariance[31]

To many neo-Kantians active in the first quarter of the 20th century, the general theory of relativity showed that Kant’s epistemology was still a work in progress, neither a refuted nor a finished edifice, and nearly all were prepared to concede, as did Hilbert, that ‘Kant greatly overestimated the role and the extent of the *a priori*’ (Hilbert 1930, 961). By the same token, it might be said that Hilbert, through the axiomatic method, was the first neo-Kantian to put his finger on where exactly 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 spatio-temporal sensory intuition of objects.

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 linkage of conditions of physical *meaningfulness* to *sensibility*, Hilbert placed general invariance as the superordinate criterion of physical objectivity, explicitly attributing this development to the influence of Einstein’s gravitational theory. This is repeatedly affirmed in his lectures, e.g., (1919–1920, 49), and (1921):

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

[30] Hilbert (1919–1920), 85–87, explicitly discusses the problem of causality in the context of general relativity, concluding (87) that “causality *(das Ursächliche)* in the narrower sense [of cause–effect relations] possesses no objective meaning for physics, and that in the search for causes, considerations of the particular conditions of human perception and of human purposes are essentially involved.”

[31] See Brading and Ryckman (2008) for a more extensive treatment of this topic.

> description of nature which is *independent* of our senses and intuition and is directed purely to the goals of objectivity and systematic unity.[32]

In broad agreement with the cognitive function that the Transcendental Dialectic assigns to reason, of seeking ever more inclusive 'systematic unity', the axiomatic method elevates the *axiom* of general invariance as the guiding principle of a 'description of nature which is *independent* of our senses and intuition'. In this sense, the principle of general invariance *is neither true nor false, but a regulative idea*,

> if, in agreement 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.[33]

Thus, for Hilbert, the axiom of general invariance is the anchor on which the objective scientific description of nature must now rest, even though such a description goes beyond the limits of possible experience, which is always finite and whose conditions are set by sensibility (representation in space and time) and the understanding (causality). By the early 1920s, Hilbert had termed his revised understanding of the *a priori* 'the finite point of view', 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:

> We see therefore: in the Kantian theory of the *a priori (Apriori-Theorie)* there is still contained anthropomorphic slag *(Schlacke)*, from which it must be freed, and after such removal only that *a priori* point of view *(apriorische Einstellung)* is left, which also lies at the foundation of pure mathematical knowledge: it is essentially that finite point of view characterized by me in different essays. (Hilbert 1930, 962)

Final Remarks

Before concluding, we offer a brief remark on the alleged 'priority dispute' over the discovery of the generally covariant gravitational field equations. Based on the account of Hilbert's aims and methods given here, it is clear that 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 as they were physical. We thus 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).'[34]

---

[32] Hilbert, 1921, *Grundgedanken der Relativitätstheorie*, lectures in SS 1921, ed. by Paul Bernays; as cited and translated in Majer (1995, 146).

[33] Hilbert (1925), Engl. trans., 392.

[34] Felix Klein (1917), 566, fn 8. '*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 footnote added to the 1921 reprint of Klein (1917).

In conclusion, we have argued that Hilbert's two notes on the Foundations of Physics can be seen as having the common goal of pinpointing, then resolving, the apparent tension between general covariance and causality, and that his approach to this issue should be understood within the logical and epistemological context of his axiomatic method. Adopting general invariance as an axiom while observing, on the basis of his Theorem I, that the initial value problem was not well posed, Hilbert even contemplated surrendering causality (in the sense of Cauchy determination). Yet when he found a way to restore causality in the face of general covariance (essentially by imposing gauge conditions), he subordinated the principle of causality—as a condition of possible experience—to general covariance, an overriding principle of physical objectivity. This achievement, along with the results displaying the connections between general covariance and features of electromagnetic theory, were obtained through application of the axiomatic method. Only in this context can we understand what Hilbert sought to do, and evaluate his success.

References

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

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

Brading, Katherine A. (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: Birkhäuser, 125–35.

Brading, Katherine A. and Thomas A. Ryckman (2008). "Hilbert's 'Foundations of Physics': Gravitation and Electromagnetism Within the Axiomatic Method", *Studies in the History and Philosophy of Modern Physics* 39, 102–53.

Corry, Leo (1999). "Hilbert and Physics (1900–1915)", in Jeremy Gray (ed.), *The Symbolic Universe: Geometry and Physics 1890–1930*. Oxford: Oxford University Press, 145–88.

——. (2004). *David Hilbert and the Axiomatization of Physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*. Dordrecht: Kluwer Academic Publishers.

Corry, Leo, Jürgen Renn, and John Stachel (1997). "Belated Decision in the Hilbert–Einstein Priority Dispute", *Science* 278, 1270–3.

Earman, John and Clark Glymour (1978). "Einstein and Hilbert: Two Months in the History of General Relativity", *Archive for History of Exact Sciences* 19, 291–308.

Einstein, Albert (1915). "Die Feldgleichungen der Gravitation". *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*, 844–847. Reprinted in A. J. Kox, Martin Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein, volume 6.* Princeton: Princeton University Press, 1996, 244–9.

——. (1916). *"Die Grundlage der allgemeinen Relativitätstheorie", Annalen der Physik* 49: 769–822. Reprinted in A.J. Kox, Martin Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein, volume 6.* Princeton: Princeton University Press, 1996, 283–339.

Ewald, William (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics.* Oxford: Clarendon Press.

Gray, Jeremy (2000). *The Hilbert Challenge.* Oxford: Oxford University Press.

Guth, E. (1970). "Contribution to the History of Einstein's Geometry as a Branch of Physics", in Moshe Carmeli *et al.* (eds.), *Relativity.* New York: Plenum Press, 161–207.

Hallett, Michael (1994). "Hilbert's Axiomatic Method and the Laws of Thought", in Alexander George (ed.), *Mathematics and Mind*. New York: Oxford University Press, 158–200.

Hilbert, David (1899). *Grundlagen der Geometrie*. In *Festschrift zur Feier der Enthullung des Gauss-Weber Denkmals in Göttingen*. Leipzig, Teubner.

——. (1900). *"Mathematische Probleme". Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris. Königlichen Gesellschaft der Wissenschaften zu Göttingen, Nachrichten,* 253–97. Translation in Jeremy Gray (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. Göttingen, SUB Cod. Ms. 634. Translation in Renn and Schemmel (2007), 989–1001.

——. (1915b). *"Die Grundlagen der Physik: Erste Mitteilung", Königliche Gesellschaft der Wissenschaften zu Göttingen. Nachrichten. Mathematische-Physikalische Klasse*, 395–407. Translation in Renn and Schemmel (2007), 1003–15.

——. (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.

——. (1917). *"Die Grundlagen der Physik: Zweite Mitteilung", Königliche Gesellschaft der Wissenschaften zu Göttingen. Nachrichten. Mathematische-Physikalische Klasse*, 53–76. Translation in Renn and Schemmel (2007), 1017–38.

——. (1918). *"Axiomatisches Denken", Mathematische Annalen*, 78, 405–15. Translated by William Ewald as "Axiomatic Thought", in Ewald (1996), 1105–15.

——. (1919–1920). *Natur und mathematisches Erkennen. Vorlesungen, gehalten 1919–1920 in Göttingen. Nach der Ausarbeitung von P. Bernays*. Published and edited by David Rowe. Basel: Birkhäuser, 1992.

——. (1922). *"Neubegründung der Mathematik. Erste Mitteilung*", *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, 1, 157–77.

Translated by William Ewald as "The New Grounding of Mathematics. First Report", in Ewald (1996), 1115–34.

——. (1924). *"Die Grundlagen der Physik", Mathematische Annalen* 92, 1–32.

——. (1925). "*Über das Unendliche", Mathematische Annalen* 95, 161–90. Translated by Stefan 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, 367–92.

——. (1930). "*Naturerkennen und Logik*", *Die Naturwissenschaften* 18, 959–63. Translation by William Ewald as "Logic and the Knowledge of Nature", in Ewald (1996), 1157–65.

Hoefer, Carl (2000). "Energy Conservation in GTR", *Studies in History and Philosophy of Modern Physics* 31, 187–99.

Klein, Felix (1917). *"Zu Hilberts erster Note über die Grundlagen der Physik", Königliche Gesellschaft der Wissenschaften zu Göttingen. Nachrichten. Mathematisch-Physikalische Klasse.* As reprinted in Felix Klein, *Gesammelte Abhandlungen Bd I.* Berlin: Julius Springer, 1921, 553–67.

Majer, Ulrich (1995). "Hilbert's Finitism and the Concept of Space", in Ulrich Majer and Heinz-Jürgen Schmidt (eds.), *Semantical Aspects of Spacetime Theories*. Mannheim: Wissenschaftsverlag, 145–57.

——. (2001). "The Axiomatic Method and the Foundations of Science: Historical Roots of Mathematical Physics in Göttingen", in M. Redei and M. Stöltzner (eds.), *John von Neumann and the Foundations of Quantum Physics*. Dordrecht: Kluwer Academic Publishers, 11–33.

Mehra, Jagdish (1974). *Einstein, Hilbert and the Theory of Gravitation.* Dordrecht: Reidel.

Mie, Gustav (1917). *"Die Einsteinsche Gravitationstheorie und das Probleme der Materie", Physikalische Zeitschrift* 18, 574–80; 596–602.

Pais, Abraham (1983). *'Subtle is the Lord... 'The Science and Life of Albert Einstein.* New York: Oxford University Press.

Pauli, Wolfgang, Jr. (1921). *Relativitätstheorie.* In Arnold Sommerfeld (ed.), *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwedungen.* Leipzig: B.G. Teubner, 539–775. Translated as *The Theory of Relativity.* Oxford: Pergamon Press, 1958.

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

Renn, Jürgen and Matthias Schemmel (eds.). (2007). *The Genesis of General Relativity*, vol. 4. *Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.* Dordrecht: Springer.

Renn, Jürgen and John 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. Reprinted in Renn and Schemmel (2007), 857–973.

Rowe, David (2001). "Einstein Meets Hilbert: At the Crossroads of Mathematics and Physics", *Physics in Perspective* 3, 379–424.

Sauer, Tilman (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 September 10, 2001.

——. (2005). "Einstein Equations and Hilbert Action: What is Missing on p.8 of the Proofs for Hilbert's First Communication on the Foundations of Physics", *Archive for History of Exact Sciences* 59, 577–90.