Invariance Principles as Regulative Ideals: From Wigner to Hilbert

THOMAS RYCKMAN

Wigner

Eugene Wigner's several general discussions of symmetry and invariance principles are among the canonical texts of contemporary philosophy of physics. Wigner spoke from a position of authority, having pioneered (and won the Nobel prize in 1963) for recognition of the importance of symmetry principles from nuclear to molecular physics. But perhaps recent commentators have not sufficiently stressed that Wigner always took care to situate the notion of *invariance principles* with respect to two others, *initial conditions* (or *events*) and *laws of nature*. Wigner's first such general consideration of invariance principles, an address presented at Einstein's 70th birthday celebration, held in Princeton on 19 March 1949, began by laying out just this distinction, and in a way that seems to suggest that the three notions arise through abstraction in an analysis of the general problem of cognition in the natural sciences:

> The world is very complicated and it is clearly impossible for the human mind to understand it completely. Man has therefore devised an artifice which permits the complicated nature of the world to be blamed on something which is called accidental and thus permits him to abstract a domain in which simple laws can be found. The complications are called initial conditions; the domain of regularities, laws of nature. (...) the underlying abstraction is probably one of the most fruitful the human mind has made. It has made the natural sciences possible.[1]

That we are able to discover laws of nature at all, Wigner went on to argue, arises from the circumstance that many regularities are relatively insensitive to initial conditions; that is, given essentially the same initial conditions, the same regularities result without regard to displacements of the system in space and time. It is here that the

[1] 'Invariance in Physical Theory', *Proceedings of the American Philosophical Society* 93, 1949; as reprinted in Wigner (1995), pp. 283–93.

doi:10.1017/S1358246108000040 © The Royal Institute of Philosophy and the contributors 2008

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

notion of an invariance principle arises in the first instance (to be sure, as a 'geometrical', not a 'dynamical', principle of invariance).

First an observation, and then a comment on Wigner's conceptual constellation of *initial conditions, laws of nature*, and *invariance principles*. An observation: the distinction between initial conditions and laws of nature, simply taken over from the employment of ordinary differential equations in classical mechanics, is long familiar in philosophy of science and is widely assumed, mistakenly perhaps, to be unproblematic. For example, it is presupposed in D-N accounts of scientific explanation and many discussions of laws of nature. Certainly, philosophical accounts widely diverge over the cognitive status of laws of nature, whether there are such, whether they need be universal, and whether they are to be analyzed in regularity or necessitarian terms. In short, the status of laws of nature remains an issue of lively contention in philosophy of science. On the other hand, philosophical examination of invariance principles is relatively recent, occurring mostly within philosophy of physics rather than philosophy of science more broadly, and has had surprisingly little impact in the debate over scientific realism.[2] This is all the more noteworthy since leading unification programs in current physics adopt the postulate that Nature at the fundamental level is *supersymmetric*, while pronouncements like 'The secret of Nature is symmetry' from highly regarded fundamental physicists have become part and parcel of even the wider intellectual culture.[3]

Secondly, a comment. It is worth emphasizing that Wigner described the tripartite distinction between *initial conditions, laws of nature*, and *invariance principles* as 'an artifice' devised by human thought when confronting the enormous complexity of natural phenomena with, he implies, only the finite capacities of the human mind. Nonetheless, this artifice is 'one of the most fruitful (abstractions) the human mind has made ... (since) it has made the natural sciences possible'. Anticipating a bit, these remarks do suggest something of a 'finitist' philosophical tendency, whose broader outlines are readily surmised from other of Wigner's essays. For one thing, it is quite clear that Wigner himself held a non-realist or antirealist attitude towards physical theories; e.g., he stated, echoing Bohr, that

> Physics doesn't describe nature. Physics describes regularities among events and *only* regularities among events.[4]

[2] van Fraassen (1990); Brading and Castellani (2003); and very recently, Debs and Redhead (2007).

[3] Gross (1995).

[4] 'Symmetry in Nature' (1973), in Wigner (1995), p. 382.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

Moreover, Wigner evinced a rather pragmatic attitude towards invariance principles, e.g., once observing that the physicist's belief in *CPT* symmetry

> is not based on nature's innate preference for symmetry; it is based on the stubborn fact that we cannot formulate equations of motion in quantum field theory that lack this symmetry and still satisfy the postulates of Einstein's theory of special relativity.[5]

Still further evidence of Wigner's dissent from realism can be found in his somewhat infamous view that resolution of the quantum measurement problem requires reference to consciousness. But rather than linger over these details, I wish, above all, to call attention to his characterization of these conceptual abstractions or patterns of understanding (*initial conditions, laws of nature*, and *invariance principles*) as a presupposition of physical theory ('It has made the natural sciences possible'). Putting matters this way suggests another, far older, attempt, rather more well-known to philosophers, to explain the possibility of the natural sciences, one also resting on a tripartite analysis of the three sources of human cognition:

> 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).

This programmatic statement, occurring at the introduction to the Transcendental Dialectic, is Kant's most concise statement of how cognition originates in the distinct contributions of intuitions, concepts, and ideas. It is repeated almost *verbatim*, presumably to underscore its significance, as a conclusion at the end of the Transcendental Dialectic

> Thus all human cognition begins with intuitions, goes from there to concepts, and ends with ideas. (A702/B730)

Though little commented upon, the latter passage, appears as the epigram of Hilbert's classic Gauss–Weber *Festschrift* essay, *Grundlagen der Geometrie* (1899). I would not want to make too much of the fact that Wigner served as Hilbert's assistant in Göttingen in 1927–28, tasked with bringing Hilbert up-to-speed on the applications of group theory to the new quantum mechanics that Wigner, with the assistance of John von Neumann, pioneered. Nonetheless, as hinted above, I think that there is something of a 'family resemblance'

[5] 'Violations of Symmetry in Physics' (1965), in Wigner (1995), p. 355.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

between Wigner's discussion and Hilbert's own physical applications of what he termed the 'axiomatic method', a method that, in published articulation, first appeared as applied to Euclidean geometry in *Grundlagen der Geometrie*.

Hilbert

The axiomatic method

At least in the scant literature on Hilbert's work in physics, it is widely assumed that Hilbert's references to 'axiomatic method' signal a typically modern mathematical concern with the rigorous explicit statement of a theory. Just so, but the 'axiomatic method' for Hilbert also connoted a specifically *logical and epistemological* method of investigation for 'deepening the foundations' of the theory. By invoking the 'axiomatic method', Hilbert called attention to a specific 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.[6]

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'.[7] Analysis next determines the concepts under which such given facts can be classified and arranged. Finally one attempts to formulate the most general relations among these concepts, a 'framework of concepts' (*Fachwerk von Begriffen*) based upon the fewest possible number of axioms. The axioms are, as far as possible, independent of the

[6] In the 1905 Summer Semester Lectures *Logische Principien des mathematischen Denkens*, Hilbert writes 'The general idea of [the axiomatic method] always lies behind any theoretical and practical thinking', as cited and translated in Michael Hallett, 'Hilbert's Axiomatic Method and the Laws of Thought', in George (1994), p. 162.

[7] David Hilbert, *Neubegründung der Mathematik. Erste Mitteilung*, Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 1 (1922), as reprinted in Hilbert (1935), pp. 157–77, 161. 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.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

particular intuitions (and so, concrete facts) with which the analysis began. By subjecting intuitively given data to such an 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. According to Hilbert, 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.[8]

As applied to any theory covering a sufficiently known (i.e., mathematized) domain of facts, whether of mathematics or natural science, the axiomatic method is a procedure for finding, for any given proposition of the theory, the premises from which it follows. 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 a given body of knowledge rests. 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. As its culmination, the axiomatic method is concerned to demonstrate that the axioms for the theory possess the three metalogical properties or relations of mutual consistency, independence, and completeness.[9] 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.[10]

[8] David Hilbert, Winter Semester lectures 1922–3 *Wissen und mathematisches Denken*. Ausgearbeitet von Wilhelm Ackermann. Mathematische Institut Göttingen. Published in a limited edition, Göttingen, 1988.

[9] Hilbert's 1905 Summer Semester Göttingen lectures, *op. cit.* note 6, already characterized the general idea of the axiomatic method as stressing the consistency, independence, and completeness of an axiom system.

[10] Hilbert (1918) as reprinted in Hilbert (1935), pp. 146–56, 148: 'The procedure of the axiomatic method (. . .) 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.'

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

Two further considerations require emphasis. First, a theory thus axiomatized 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. Yet the axiomatic method requires that consistency obtain not only with respect to the various axioms, but also with respect to what Hilbert termed 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'.[11] This emphasis on the compatibility between the different sources of knowledge is crucial to understanding Hilbert's project in relativistic physics in 1915–17.

Secondly, the mathematical axioms standing at the pinnacle of the *Fachwerk von Begriffen* are not only general but also *ideal*: as we shall see, Hilbert regarded them as *Ideen* in Kant's regulative sense, principles that have a guiding regulative significance but not a constitutive employment in cognition (however, Hilbert's revision of Kantian epistemology rejects a sharp *constitutive/regulative* distinction). In virtue of their ideality, and so severance from experience and intuition, axioms can play at best a hypothetical or heuristic role in cognition.

Hilbert's application of the axiomatic method in 1915–17

As many will know, in November 1915, Einstein and Hilbert were engaged in what might be described as a sprint to the finish.

[11] 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 '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), p. 274.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

Many will also know that the 'priority question' regarding the generally covariant gravitational field equations recently has been revisited with an unequivocal decision in Einstein's favor. Moreover, Hilbert's two papers on generally invariant physics of 1915 and 1917, both then, and ever since, have been assessed solely in terms of the contributions they made, or failed to make, to general relativity. Now *our* view (here I am drawing on recent work carried out jointly with Katherine Brading)[12] is that Hilbert's work merits examination on its own terms, namely, as an explicit treatment of gravitation and electromagnetism within the framework of the 'axiomatic method' where the principle of general covariance (Hilbert: general invariance) is adopted as an *axiom*. The issue of 'priority' is here an enormous distraction, preventing any detailed internal examination.[13]

First, the facts. In November and December 1915, Hilbert made two presentations to the Göttingen Academy of Sciences under the common title *Die Grundlagen der Physik* but distinguished as 'First Contribution' and 'Second Contribution'. Two papers with these designations eventually appeared in the *Nachrichten* of that Academy, the latter not until January 1917. In his initial communication, on 21 November 1915, Hilbert sketched an elegant variational argument, first constructing what was subsequently termed the 'Hilbert–Einstein Action' based on the Riemann scalar, from which follow, under certain further assumptions, the generally covariant field equations of gravitation. As we know, in November, Einstein submitted four papers to the Berlin Academy: one to each weekly meeting. The last of these, on 25 November 1915, contained, for the first time, the generally covariant field equations of gravitation known to the world since as the Einstein Equations. However, there are *two* extant versions of Hilbert's first contribution: one eventually published in the *Nachrichten* on 31 March 1916 (the delay presumably due to a wartime paper shortage) as well as the printer's proofs stamped with the date 6 December 1915 that was found by Leo Corry in 1996 in archives at the Mathematical Institute in Göttingen. As was widely reported at the time, Corry's discovery

[12] Brading and Ryckman (2008).

[13] In our defense of this mode of proceeding, there is the assessment of Felix Klein, who noted, in 1921: '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)'. Klein (1917); as reprinted with additional remarks in Klein (1921), p. 566, fn 8.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

of this document brought a 'belated decision' to the priority question.[14] For the Einstein field equations do not explicitly appear in the printer proofs of Hilbert's paper but only as Euler–Lagrange derivatives, whereas they do so explicitly appear in the published version. Einstein's 25 November 1915 presentation had been published long before, on 2 December 1915, in the *Sitzungsberichte* of the Berlin Academy.

These new developments obviously raise troubling questions regarding a seeming gap between Hilbert's claims and what he actually accomplished in November 1915. Katherine Brading and I believe we have shown that Hilbert's aims and his achievements only in part overlap with Einstein's, but above all must be seen as deeply embedded within the epistemological context of the 'axiomatic method'. The purpose of the remainder of this paper is to briefly present reasons why one might think this, while showing that Hilbert's far more explicit treatment of axioms of invariance in physical theory as 'regulative ideals' conceivably elaborates and amplifies Wigner's more guarded thoughts on the role and significance of invariance principles.

Notably, at the 1900 International Congress of Mathematicians in Paris, Hilbert had proposed 'the axiomatisation of physics' as the sixth of his famous twenty three 'mathematical problems'.[15] Both versions of his first contribution on 'Foundations of Physics' are Hilbert's attempt to solve that problem, subsuming all known physical interactions in 1915 (i.e., gravitation and electromagnetism) under two axioms that essentially present the fundamental laws of physics as solutions to a variational ('least action') problem, under the guiding condition of general invariance:

AXIOM I ('Mie's Axiom of the World Function'). Using the term 'world function', familiar from his lectures on axiomatizing physical theories since 1905, Hilbert formulated a variational argument for a 'world function' H, 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:

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

[14] Corry, Renn, and Stachel (1997).

[15] See Gray (2000) for discussion and an English translation of Hilbert's 1900 lecture.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

AXIOM II ('Axiom of General Invariance'). The world function H is an invariant with respect to arbitrary transformations of the 'world parameters' w_s $(s = 1, 2, 3, 4)$.[16] 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 . . . completely independent of the way in which one designates the world points through world parameters.[17]

Hilbert next turns to the derivation of the Euler–Lagrange differential equations for gravitation and electromagnetism. The former arise from differentiation of H with respect to the gravitational potentials $g_{\mu\nu}$ and their first and second derivatives, yielding (using Hilbert's numbering) ten gravitational equations,

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

or, in Hilbert's abbreviation, (4)

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

The latter arise from differentiation of H with respect to the electromagnetic potentials q_s and their first derivatives, yielding four equations of electrodynamics, which, following Mie as Hilbert

[16] Hilbert's use of the term 'world parameters' in place of the standard locution 'space-time coordinates' is instructive. As expressly stated in his (1917), and as Mie also observed, 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.

[17] Hilbert (1915a). Also (1915b), p. 396.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

indeed did, is the basis of Mie's theory of matter:

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

$$\text{or, } [\sqrt{g}H]_h = 0 \text{ [notation: } q_{hk} = \frac{\partial q_h}{\partial w_k}(h, k = 1, 2, 3, 4)].^{18}$$

The fourteen equations 4 and 5 are termed respectively, 'the basic equations of gravitation' and 'the basic equations of electrodynamics or generalized Maxwell equations'. On the assumption that the Mie electromagnetic theory of matter was viable, these equations encompassed the entirety of known fundamental physics in 1915.

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

THEOREM I ('*Leitmotiv*'). In the system of n Euler–Lagrange differential equations in n variables obtained from a generally covariant variational integral such as in Axiom I, 4 of the n equations are always a consequence of the other $n-4$ in the sense that 4 linearly independent combinations of the n equations and their total derivatives are always identically satisfied.[19]

[18] The form of equations 4 and 5 is trivially algebraically different between (1915a) and (1915b). Here the published version (1915b) is followed. For ease of comparison with the text, 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' is followed.

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

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

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 (1918). However in a passage appearing in the December Proofs, but removed 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.[20]

As befits its preeminent concern with the *consistency* of all axioms and assumptions that underlie a theory, the axiomatic method has here revealed a seeming conflict between general covariance and causality in the sense of an apparent failure of univocal determination, a conflict characterized as whether *any* theory satisfying Axioms I and II admits of a well-posed Cauchy problem. Theorem I suggests that it is a property of any such theory that it does not. For a system of second order partial differential equations, the Cauchy problem is that of showing that, from given initial data assignments to the 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'). In field theories formulated on spacetime, the initial data are formulated on a given 3-dimensional spatial surface (a 'space-like hypersurface'), and the essential problem is that of showing that the field equations determine the evolution of the data off that surface according to the second time derivatives of the given field quantities. As Hilbert repeatedly emphasized in period lectures, in all physical theories prior to general relativity, Cauchy determination is satisfied in that there are precisely as many independent equations as there are independent functions to be determined. However, the situation is complicated in a generally covariant spacetime theory (as was notoriously shown by Einstein in the so-called 'Hole Argument' in 1913–5)[21] by the

[20] Hilbert (1915a), pp. 3–4.

[21] See e.g., Stachel, 'The Meaning of General Covariance: The Hole Story', in Earman *et al*. (eds) (1993), pp. 129–60.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

freedom to make arbitrary coordinate transformations (equivalently, local diffeomorphic point transformations) of solutions to the field equations. In terms of Theorem I, this results from the fact that, on account of the axiom of general invariance, 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 postulate of general covariance (or, general invariance) and the requirement of causality in the mathematical sense of univocal determination*.

Notice that univocal causal determination—in the sense of Cauchy determination (one equation for each independent variable)—is not an axiom in Hilbert's construction. Nevertheless, it is a requirement satisfied by previous physical theories, and so its seeming failure in the context of general covariance surely sparked Hilbert's interest. In short, a central outcome Hilbert reached by means of the axiomatic method is this: *any generally covariant spacetime theory raises deep questions about causality, in both the mathematical and the physical sense* (*and in the distinction between these*).

Accordingly, Hilbert set himself the task, in the December Proofs, to find the four missing equations required for a well-posed Cauchy problem. After the lengthy construction of an 'energy form', the requisite four equations are derived by imposing special coordinate restrictions on an 'energy vector', namely that it must satisfy the divergence equation, given again in Hilbert's numbering,

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0, \text{ iff the four quantities } e_s = 0. \qquad (15)$$

Here the w_i are now special spacetime coordinates adapted to this 'energy theorem', 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

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

narrowest connection between the energy theorem and the principle of causality, in that each conditions the other.[22]

Now 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 1913–4, where four non-generally covariant equations ensure energy conservation by restricting the covariance class of the field equations.[23] But Hilbert's rather more complicated construction has, philosophically and motivationally, a different *raison d'être*. His 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 invariance, which is, after all, an *axiom*).

The very complex mathematical derivation in the Proofs leading to the four non-covariant energy equations was cut, together with *all* of its motivation, from the published version. The reason, surely, is that Einstein's 25 November 1915 paper contained an implicitly generally covariant treatment of energy conservation. So Hilbert recognized he had adopted the wrong approach for resolving the tension between general covariance and Cauchy-determination. In the published version, Hilbert derived a generally covariant 'energy equation' that anyway is consonant with the so-called 'trace' term in the gravitational field equations popping out through explicit calculation from their Lagrangian derivatives. But then, *also in the published version, all reference to the tension between general invariance and causality implied by Theorem I was deleted*!

Of course, the issue of causality in a generally covariant theory did not go away.[24] Hilbert's second contribution[25] indeed begins with a discussion of the problem of causal order, and contains a much revised, and lengthy, reconsideration of this issue. It is rightly understood as resolving the problems posed in the December Proofs.

Without going into much further detail, which is largely irrelevant here, the strategy in this second Hilbert paper is to adopt a criterion of physical meaningfulness and objective validity in terms of general

[22] Hilbert (1915a), 7.

[23] Renn and Stachel (1999), p. 32, note Einstein's conviction '(e)ven before Einstein developed the hole argument', that energy–momentum conservation requires such a restriction.

[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, and remains a delicate issue.

[25] Hilbert (1917).

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

invariance, while recognizing that human experience (measurement and observation) requires a restricted representation of space and time in which the causal relation is preserved.[26] To this end, he introduced what he called 'proper coordinate systems'.[27] In the case of a general Riemannian metric, these amount to so-called 'reality relations'[28] reflecting the physical requirement of metrical indefiniteness: that 3 of the coordinate axes are space-like, and 1 time-like. Moreover, since a 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', ensuring that causal order between events along such a curve is always preserved. This reconciliation, set in the frame of the Cauchy problem, produced the four needed coordinate conditions to restore univocal determination of future states from present ones, and the requirement of causality is accordingly modified:

> 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 (31) in addition to our fundamental equations; that is, we restrict ourselves to the use of *proper* space-time coordinates.[29]

[26] In brief, general invariance (or, general covariance) allows for local diffeomorphic point-transformations among the solutions to the field equations; for causal (time-like) curves, this permits inversion of the ordering of space-time points on such curves. Obviously, this can lead to causal anomalies. Hilbert first recognized that the full invariance (or covariance) of a generally invariant theory must be broken in order to adequately establish causality (in terms of the initial value and Cauchy problems). On the other hand, Hilbert held that the fundamental equations of physics must be time-reversal invariant, and so favoring the causal ordering by adopting particular coordinate systems is but an 'anthropomorphic' epistemic condition.

[27] 'Proper coordinate systems' are defined by the satisfaction of 4 inequalities among the components of the metric tensor. Where x_4 is the time coordinate, using Hilbert's numbering (31),

$$\text{these are: } g_{11} > 0, \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0, \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0, \; g_{44} < 0.$$

[28] Pauli termed them *'Wirklichkeitsverhältniße'* in Pauli (1921), English translation (1958), §22.

[29] Hilbert (1917), p. 58. For a definition of the aforementioned inequalities (31), see footnote 27.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

Hilbert is essentially stating here that it is the structure of our cognitive experience (in Kantian terms, that imposed by our sensibility and understanding) that leads to the requirement that we use proper coordinate systems—this restriction has to do *not* with the ideal generally invariant objects of physics, 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 spatiotemporal empirical intuition.

With this conception of the causality principle in hand, the necessary and sufficient conditions for a proposition to be physically meaningful are now these:

(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 space-like past hypersurface, abiding the human epistemic interest in causal relations.

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 cognition ('physical meaningfulness'), a necessary condition imposed by the mind in structuring experience. Like the subjectivity of the sense qualities, the requirement of physical causality is *anthropomorphic*, having to do not with the objective world of physics but rather with our experience of that world.

It might be said that Hilbert, through the axiomatic method, was the first to put a finger on where exactly the general theory of relativity required a modification of the traditional Kantian transcendental framework that expressly binds considerations of objectivity to conditions of possible experience. According to Kant, for a cognition to be *objectively valid* (i.e., to be a representation pertaining to a possible object *for us*, hence to be *meaningful*) is for it to invoke *our* specifically human type of finite, receptive spatiotemporal sensory intuition of objects. Hilbert essentially argued that this is no longer the case once the requirement of general invariance is adopted in 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.

The 'axiomatic method' is ideally suited to show this, for it is the instrument for working up a domain of intuitively presented physical facts into a purely logical-formal system of relations, crowned in

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

Hilbert's papers of 1915 and 1917 by an axiom positing a generally invariant spacetime integral. However, 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, away 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 requires a (never fully attainable) emancipation 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 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.[30]

Fundamental physical theory, in accordance with the axiom of general invariance aspires to a *non-anthropomorphic* account of physical objectivity, that is, to a 'description of nature which is *independent* of our senses and intuition'. But 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.[31]

[30] Hilbert, *Grundgedanken der Relativitätstheorie*, Göttingen Summer Semester Lectures 1921; as cited by Majer (1995), p. 284.

[31] Hilbert (1926), p. 190.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

As Ulrich Majer in particular has emphasized, by employing the '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.[32]And, while Hilbert allowed that such elements could be introduced either through mathematical construction or as axioms,[33] the axiomatic method, naturally, prefers the latter. In virtue of their ideality, and so severance from experience and intuition, axioms of invariance play a hypothetical role in cognition, no longer constitutive of experience, but informing and guiding the idea of physical objectivity.

This completes Hilbert's revision of the Kantian a priori. By the early 1920s, Hilbert had termed his revised understanding of the a priori 'the finite point of view' (*die finite Einstellung*). But Hilbert's work on 'Foundations of Physics' in 1915–17 shows in addition that Majer is correct in concluding that

> Hilbert's 'finite point of view' is not restricted to mathematics or meta-mathematics, but is stated as a *universal principle of epistemology*.[34]

Hilbert took from Kant the methodology or standpoint that objective cognition can only be understood as conditioned by a priori structures of the mind, but he refashioned 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

[32] Majer (1993a).

[33] E.g., Hilbert (1992), p. 91: 'The transition to the broader system [containing ideal elements] either can be effected in a constructive manner, in which the new elements are developed through mathematical construction from the older ones, or in an axiomatic manner, in which the new system is characterized through relational properties. In the second case, it requires a proof that the supposition of a system with the desired condition does not in itself lead to contradiction.'

[34] Majer (1993b), p. 191.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press

> (*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.[35]

In sum, Hilbert's two communications on 'Foundations of Physics' are a complex epistemological response to generally covariant field physics, hitherto unrecognized, and directly tied to his employment of the 'axiomatic method'. In accordance with that method, Hilbert understood the significance of the principle of general covariance within the broad framework of the Transcendental Dialectic. A product of the faculty of pure theoretical reason in its 'hypothetical' employment, the axiom of general invariance posits a regulative ideal of systematic unity, and possesses an objective but indeterminate validity. Hilbert's setting of gravitation and electromagnetism within the frame of the 'axiomatic method' is a significant example illustrating how invariance principles may be methodologically interpreted as the projection of a necessary postulate of the systematization or unity of nature grounded in the basic structure of human reason and cognition. It would seem from Wigner's remarks on invariance principles that something along these lines is also what he had in mind.

[35] Hilbert (1930); reprinted in Hilbert (1935), pp. 383, 385.

https://doi.org/10.1017/S1358246108000040 Published online by Cambridge University Press