JOHN STACHEL

THE STORY OF NEWSTEIN OR: IS GRAVITY JUST ANOTHER PRETTY FORCE?

1. 1. INTRODUCTION

In this paper I will argue for the following three theses:

1. The concepts of parallel displacement in Riemannian geometry and of a non-metrical affine connection were developed postmaturely (see Section 2): By the latter third of the nineteenth century, all of the mathematical prerequisites for their introduction were available, and it is a historical accident that they were not developed before the second decade of the twentieth century (see Section 3).
2. The appropriate mathematical context for implementing the equivalence principle is the theory of affine connections on the category of frame bundles, with the bundle morphisms induced by diffeomorphisms on the base manifold (see the Appendix).[1] This theory allows a mathematically precise formulation of Einstein's insight that gravitation and inertia are "essentially the same [*wesensgleich*]" as he put it (see Section 5). The absence of this context constituted a serious obstacle to the development of the general theory of relativity—indeed an insurmountable one to its development by the mathematically most direct route. Consequently, Einstein was forced to take a detour through a long and indirect route from the initial formulation of the equivalence principle in 1907 to the final formulation of the field equations in 1915 (see Section 10). The detour involved focusing attention almost exclusively on the chrono-geometrical structure of space-time, and to this day, many discussions of the interpretation of the general theory, and of the problem of quantum gravity, still reflect the negative consequences of this detour.
3. Had the concept of an affine connection been developed in a timely manner, the affine formulation of Newtonian gravitation theory, which was actually developed only *after* the formulation of *general* relativity,[2] could have been developed *before* the formulation of *special* relativity. From the outset, such a formulation would have placed appropriate emphasis on the inertio-gravitational structure of space-time and posed the question of its relation to the chronometry and geometry

1 Insofar as needed for this paper, these concepts are briefly explained in the Appendix. A particularly useful reference for a more extended discussion of most of these concepts is (Crampin and Pirani 1986).

2 See (Cartan 1923; Friedrichs 1927). Excerpts from Cartan can be found in this volume.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Theories of Gravitation in the Twilight of Classical Physics:
The Promise of Mathematics and the Dream of a Unified Theory
© 2006 Springer. Printed in the Netherlands.

of space-time (see Sections 6 and 7). When special relativity, with its new chrono-geometry, was developed, this context for gravitation theory would have made the transition from the special to the general theory of relativity rather transparent, thereby avoiding the negative consequences of the actual transition mentioned above.

In order to vivify these rather abstract theses, I have created Isaac Albert Newstein (= Newton + Einstein), a mythical physicist who combines Newton's approach to the kinematical structure of space and time (chronometry and geometry) with Einstein's insight into the implications of the equivalence principle for (Newtonian) gravitation theory (see Section 7). He did this shortly after Hermann Weylmann (= Weyl+ Grassmann), an equally mythical mathematician, formulated the concept of affine connection around 1880. Of course, Newstein had to adopt a four-dimensional treatment of space and time in order to carry out his reformulation of Newtonian gravitation theory; but, long before that, the concept of time as a fourth dimension had been introduced in analytical mechanics by d'Alembert and Langrange.[3]

Continuing my mythical account, when in 1907 Einstein turned to the problem of extending his original (later called special) theory of relativity to include gravitation, Newstein had already shown how to describe the inertio-gravitational field by a non-flat affine connection. Einstein's problem was to combine this insight about the nature of gravitation with the new chrono-geometrical structure of space-time that he had introduced in 1905. Once the problem is posed in this way, the step from Newstein's formulation of the gravitational field equations to the corresponding equations of Einstein's general relativity is a short one (see Section 8).

Of course, all of this is pure fable; but I believe that—in addition to their entertainment value—such scientific fables are of real value for the history and philosophy of science. First of all, they help us to combat the impression of inevitability often attached to the actual course of historical development, the idea that the "discovery" of a theory is just that: the bringing to light by the intellect of some pre-existing structure, previously hidden but predestined to emerge sooner or later and enter into the scientific corpus in just the form in which it actually did. Secondly, they help us to question the thesis that the formulation of a theory is more-or-less independent of its mode of discovery, including the peculiarities of the individual(s) who happened to "discover" it and the process of negotiation that led to its assimilation into the body of accepted knowledge by the scientific community. Such questions can lead to a more critical re-examination of the current formulation(s) of the theory. We are bound to look more critically at what actually happened, and at the accepted formulation(s) of a theory, if we can produce one or more credible scenarios showing how things might have happened quite differently.[4]

---

3 This is no myth. See my article on "Space-Time," in (Stachel Forthcoming).

4 See (Stachel 1994a) and, for other examples from the history of relativity, (Stachel 1995). For some further comments on alternative histories, see the final section, "Acknowledgements and a Critical Comment."

2. POSTMATURE CONCEPTS AND THE ROLE OF ABSENCE IN HISTORY

Zuckerman and Lederberg have suggested that, just as there are premature discoveries, "there are postmature discoveries, those which are judged retrospectively to have been 'delayed'" (Zuckerman and Lederberg 1986, 629).[5] I wish to apply the concept of postmaturity to theoretical entities; but since, as noted above, the word "discovery" might suggest a Platonist attitude to mathematical and physical concepts, I shall use more epistemologically neutral phrases: "postmature development," "postmature concept," "postmature theory," etc.

As the work of Zuckerman and Lederberg suggests, in retrospect one can see that—like other forms of absence—the absence of a postmature concept can play a crucial role in the dialectical interplay that shapes the actual course of historical development. My use of word "dialectical" here is purposeful. The second chapter of Roy Bhaskar's book on dialectics (Bhaskar 1993)[6] is entitled: "Dialectic: The Logic of Absences." He equates *absence* with what he calls *real negation*, whose "primary meaning is real determinate absence or non-being (i.e., including non-existence" (Bhaskar 1993, 5). He describes real negation as:

> the central category of dialectic, whether conceived as argument, change or the augmentation of (or aspiration to) freedom, which depends upon the identification and elimination of mistakes, states of affairs and constraints, or more generally ills—argued to be absences alike (Bhaskar 1993, 393).

Elsewhere I shall argue for this viewpoint with examples drawn from the history of music as well as the history of science. But to return to the central concern of this paper, my claim is that "affine connection" is a postmature concept, the absence of which during the course of development of the general theory of relativity had a crucial negative influence on its development and subsequent interpretation. Conversely, the filling of that absence opened the way to a deeper understanding of the nature of gravitation and of its relation to other gauge field theories of physics.

3. A LITTLE HISTORY

Gauss first developed the theory of curved surfaces embedded in Euclidean three-space, including the concepts of intrinsic (or Gaussian) and extrinsic curvature. But he defined these concepts in a way that did not depend on the concept of parallelism.[7]

---

5 I am indebted to Gerald Holton for drawing my attention to this paper, which fills a gap in my earlier presentations of Newstein's story.

6 I regard Bhaskar's work on critical realism as the most significant attempt at a modern Marxist approach to the philosophy of science (see Stachel 2003a). For a critical introduction to Bhaskar's work, see (Collier 1994).

7 Essentially, he defined the intrinsic curvature at a point of a surface in a way that seemed to depend on the embedding of the surface—in terms of the radius of curvature of the sphere that best fits the surface at the point in question—and then proved that the result really does not depend on the embedding. See (Gauss 1902), and for a modern discussion (Coolidge 1940, Book III, chap. III, 355–387).

The development of differential geometry had proceeded quite far by the time Riemann introduced the concept of a locally Euclidean manifold with curvature varying from point to point in 1854, first published in (Riemann 1868).[8] So the idea of starting with a geometrical structure defined in the infinitesimal neighborhood of a point of a manifold and proceeding from the local to the global structure was quite familiar by the last third of the nineteenth century.

Similarly, discussions of the concept of parallelism had played a central role in the development of non-Euclidean geometry in the first half of the nineteenth century.[9] Grassmann's work on affine geometry had abstracted the concepts of parallel lines, plane elements, etc., from their original three-dimensional, Euclidean contexts.[10] Few were aware of the first (1844) edition of the *Ausdehnungslehre*, or even of the second version in 1862; but after the publication of the second edition of the 1844 version in 1878, knowledge of his work began to spread among mathematicians, so that it was widely available to them by the last two decades of the century.[11] By this time, there was already a rich literature on the geometrical interpretation of the principles of mechanics for systems with n-degrees of freedom based on n-*dimensional Riemannian geometry*.[12]

In all this time no one applied Riemann's approach to intervals to the concept of parallelism. Karin Reich has drawn attention to the problem of the delay in the extension of the local approach in geometry to the concept of parallelism:

> Parallelism was and is thus a central theme for the foundations of geometry. Yet it is missing in Bernhard Riemann's Habilitation Lecture "On the Foundational Hypotheses of Geometry," indeed the word parallel does not occur here. Also in the succeeding period of rapidly occurring development of Riemannian geometry parallelism was not a theme. Perhaps this is one of the reasons why Riemannian geometry was not unconditionally accepted by pure geometers (Reich 1992, 78–79).[13]

---

8 For the history of differential geometry, see (Struik 1933; Coolidge 1940; Laptev and Rozenfel'd 1996, sec. 1: "Analytic and Differential Geometry," 3–26).

9 For the standard older historical-critical account of non-Euclidean geometry, see (Bonola 1955).

10 See (Grassman 1844; 1862; 1878), and for an English translation, (Grassmann 1995). For a survey of publications using Grassmann's approach, demonstrating that their number increased considerably after 1880, see (Crowe 1994, chap. 4); by the end of the century, interest in Grassmann's work was comparable to that in Hamilton's. Weyl was well aware of Grassmann's work. Speaking of affine geometry, he says: "For the systematic treatment of affine geometry with abstraction from the special 3-dimensional case, Grassmann's "Lineale Ausdehnungslehre" (Grassmann 1844)... is the groundbreaking work" (Weyl 1923, 325). In a recent discussion of Grassmann's role as a forerunner of category theory, Lawvere (Lawvere 1996) speaks of "the category A of affine-linear spaces and maps" as "a monument to Grassmann" (p. 255).

11 For a study of Grassmann and his influence, see (Schubring 1996).

12 See (Lützen 1995a; 1995b) for surveys of some of this work.

13 Readers of this work will realize the extent of my indebtedness to Karen Reich's work. I also gratefully acknowledge several helpful discussions with Dr. Reich.

Her retrospective critical judgement a century later is borne out by the contemporary evaluations of those who filled that gap in 1916–1917: Hessenberg, Levi-Civita, Weyl and Schouten.

Hessenberg's paper (Hessenberg 1917) was actually the first such paper, dated June 1916. It starts with a reference to relativity: "Because of the significance that the theory of quadratic differential forms has recently attained for the theory of relativity, the question of whether and how the elaborate and difficult formal apparatus of this theory can be simplified, if not bypassed, gains new significance (p. 187)." Speaking of "Christoffel's well-known transformational calculus," Hessenberg states that his aim is to "replace [it] with a geometrical argument (p. 187)." He criticizes the "formal methods of formation" of various quantities that occur because they do not bring out "the essentially *intuitive* [*anschaulich*] *meaning* of the invariants and covariants needed for the geometrical and physical applications" (p. 191). He stresses the role of Grassmann. "Access [to their geometrical significance] is opened in a way that, to me, seems surprisingly simple by means of Grassmann's ideas" (p. 192).

Levi Civita's paper (Levi Civita 1916), which is dated November 1916, also starts with a reference to Einstein's work:

> Einstein's theory of gravitation ... regards the geometrical structure of space ... as depending on the physical phenomena that take place in it ... The mathematical development of Einstein's magnificent conception ... involves as an essential element the curvature of a certain four-dimensional manifold and the related Riemann symbols [i.e., the curvature tensor] ... Working with these symbols in questions of such great general interest has led me to investigate if it is not possible to simplify somewhat the formal apparatus that is usually used to introduce them and to establish their covariant behavior. Such an improvement is indeed possible ... [This work] started with that sole objective, which little by little grew in order to make room for the geometrical interpretation [of the Riemannian curvature]. At the beginning I thought to have found it in the original work of Riemann ... ; but it is there only in embryo. ... [O]ne gets the impression that Riemann really had in mind that intrinsic and invariant characterization of the curvature, which will be made precise here. On the other hand, however, there is not a trace, either in Riemann or in Weber's commentary, of those specifications (the concept of parallel directions in an arbitrary manifold and consideration of an infinitesimal geodesic quadrilateral with two parallel sides) that we recognize to be indispensable from the geometrical point of view (pp. 173–174).

Reich comments:

> With this word "indispensible" Levi-Civitá recalled Luigi Bianchi's characterization of Ricci's absolute differential calculus. Bianchi had characterized this in 1901 as "useful but not indispensable" (Reich 1992, 79–80).

Weyl (1918b) states:

> The later work of Levi-Cività [1916], Hessenberg [1917], and the author [Weyl 1918a][14] shows quite plainly that the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector.[15]

14 For a discussion of this and the succeeding editions of Weyl's book, see the next section.

After the introduction of Riemannian parallelism by Hessenberg and Levi-Civita (and, again independently in (Schouten 1918)), it was but a brief and natural step to its generalization. Since the abstraction (in the large) of affine parallelism from parallelism in Euclidean geometry had already been made, the abstraction (in the small) of affine parallelism from parallelism in a Riemannian manifold is immediately suggested by the analogy. Indeed, Weyl took that step just a year later: In (Weyl 1918a) he defines an affinely connected manifold.[16]

The evidence thus indicates that both the Riemannian concept of parallelism and its affine generalization were introduced *postmaturely*. The absent concept of Riemannian parallelism could have been filled at any time during the last third of the nineteenth century, and followed quickly by the introduction of the concept of an affinely connected manifold, since it is a natural generalization of the Grassmannian "*lineale Ausdehnungslehre*."

Indeed, Grassmann himself might have accomplished these tasks. Towards the end of his life he learned about the work of Riemann and Helmholtz, and one of his last publications (Grassmann 1877) discusses the relation of their work to his *Ausdehnungslehre*. He discusses a method of introducing such non-linear geometries that amounts essentially to defining them as subspaces of linear spaces of higher dimensions. The path that Levi-Civita initially took to the definition of Riemannian parallelism was based on embedding a Riemannian space in a Euclidean space of sufficiently high dimension. Had Grassmann lived longer, it is conceivable that he might have introduced the concept of affine parallelism by a similar method (see the discussion in the Appendix). But he died in the same year that he wrote this paper; so I have been forced to invent Weylmann, the mathematician who introduces the concept of an affinely connected manifold around 1880, neither prematurely nor postmaturely.

4. EQUIVALENCE PRINCIPLE AND AFFINE CONNECTION

It was Albert Einstein who first realized the profound significance of the equality of inertial and gravitational mass. He soon began to speak of inertia and gravitation as "*wesensgleich*": essentially the same in nature. By an acceleration of the frame of reference, the division between inertial and gravitational "forces" can be altered, and indeed by a suitably chosen acceleration the combination of both can even be made to vanish at any point of space-time.

Einstein's problem was to find the way to incorporate this physical insight into the mathematical structure of gravitation theory. After the development of the concept of affine connection, the way became clear: there is an inertio-gravitational field, repre-

15 Translated from (Weyl 1923, 202).

16 For references and discussion of the work of Levi-Civita, Hessenberg, Schouten and Weyl, see the indispensible (Reich 1992). For the background to Weyl's "Purely Infinitesimal Geometry," see (Scholz 1995). I am indebted to Dr. Erhard Scholz for a discussion of this work.

sented mathematically by a symmetric connection in space-time, which incorporates this essential unity in its very nature. We can see the development of this insight by looking at the various editions of Weyl's *Raum-Zeit-Materie*. In (Weyl 1918b), Levi-Civita's concept of parallel transport, based upon the embedding of a Riemann space in a flat Euclidean space of higher dimension, is freed from this dependence by giving it an intrinsic definition. Weyl further states that the Christoffel symbols represent the gravitational field. In (Weyl 1919)—which follows the argument of Weyl (1918a)—the concept of parallel transport is freed from its dependence on the metric field by the introduction of the concept of affine connection. Weyl (1921) refers to this connection as the "guiding field" (*Führungsfeld*), incorporating the effects of both gravitation and inertia on the motion of bodies.

Soon afterwards, Cartan (1923) drew the obvious conclusion: By incorporating the equivalence of gravitation and inertia into Newton's gravitation theory, it can be formulated in terms of a Newtonian affine connection. Since then, starting with (Friedrichs 1927) and culminating—but certainly not ending—in (Ehlers 1981), a series of refinements of Cartan's approach have brought the affine version of Newton's theory to a state of considerable mathematical perfection.

However, I shall not give the most, abstract, coordinate-free characterization of the Newtonian affine connection based on the simplest set of axioms. For our purposes, it will be more useful to show how, starting from the usual form of the Newtonian theory of gravitation, the components of the connection with respect to a physically chosen basis may be defined, thus suggesting how Newstein could have proceeded—had he only existed![17]

5. NEWSTEIN'S WORLD

We shall start from the usual formulation of Newtonian gravitation theory in some inertial frame of reference (ifr, for short). Events in this frame are individuated with the help of the Newtonian absolute time t (chronometry), and three Cartesian coordinates (i.e, assuming Euclidean geometry), fixed relative to some choice of origin O and of three mutually perpendicular axes.[18] Since inertial and gravitational mass are equal, if $\boldsymbol{g}$ represents the force/unit gravitational mass, the equation of motion of a (structureless) particle will be

$$\boldsymbol{a} = \boldsymbol{g}, \tag{1}$$

where $\boldsymbol{a}$ is the acceleration of the particle with respect to the chosen ifr.

---

17 See (Stachel 1994b) for a somewhat more abstract discussion of space-time structures in Newton-Galilean and special-relativistic space-times (i.e., in the absence of gravity), and in Newtonian and Einsteinian gravitational theories.

18 We assume units of time and distance fixed initially and used in all frames of reference, and shall use vector notation, so that, for example, the displacement vector from the origin $\boldsymbol{r} = (x^1, x^2, x^3)$, the velocity $\boldsymbol{v} = d\boldsymbol{r}/dt$, the acceleration $\boldsymbol{a} = d\boldsymbol{v}/dt$, etc, all with respect to the ifr, are denoted by boldface symbols.

Now consider transformations to another frame of reference, moving linearly with respect to the first:

$$\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{R}(t), \quad t' = t. \tag{2}$$

If the velocity vector $\boldsymbol{V} = d\boldsymbol{R}/dt$ is *constant*, then the transformation is to another inertial frame of reference, and the equation of motion, eq. (1), is invariant under such a transformation. That is, both $\boldsymbol{a}$ and $\boldsymbol{g}$ are invariant under such *Galilei transformations* from one inertial frame to another.

However, if $\boldsymbol{V}$ is *not* constant, then the transformation is to some linearly accelerated (rigid) frame of reference, and differentiation of eq. (2) twice with respect to the time gives

$$\boldsymbol{a}' = \boldsymbol{a} - \boldsymbol{A}(t), \quad \boldsymbol{A}(t) = d^2\boldsymbol{R}(t)/dt^2. \tag{3}$$

In Newtonian mechanics, "true" forces, such as $\boldsymbol{g}$, are assumed to be the same in all frames of reference. To compensate for the use of a non-inertial frame of reference, so-called "inertial forces" appear in the equations of motion (such forces might better be called "non-inertial"). Indeed, when we substitute eq. (3) in eq. (1), we get:

$$\boldsymbol{a}' + \boldsymbol{A} = \boldsymbol{g}, \text{or } \boldsymbol{a}' = \boldsymbol{g} - \boldsymbol{A}, \tag{4}$$

and $-\boldsymbol{A}(t)$ appears as such an "inertial force" in the equation of motion of a particle with respect to a linearly accelerating frame.

But, one may ask, if we carry out measurements in some frame of reference, and get an acceleration, let us say $\boldsymbol{a}'$, for a test particle, how do we separate it into its components, the "true force" $\boldsymbol{g}$ and the "inertial force" $-\boldsymbol{A}$? Newton would not have hesitated a moment in answering: Look for the sources of the gravitational force, and use the inverse square law to compute the total $\boldsymbol{g}$ at the point where the test particle is located. Alternatively, he might have proposed: Look at the center of mass of the "system of the world" (i.e., the solar system) and see whether you are accelerating relative to it to find $\boldsymbol{A}$.

But by the end of the nineteenth century, under the influence of Maxwell's electromagnetic theory, the field point of view towards forces was beginning to prevail; according to this viewpoint, one should look upon the gravitational field as the conveyor of all gravitational interactions between massive bodies. Accordingly, the local gravitational field at a point in space (and an instant of time) should always be ascertainable by means of local measurements in the neighborhood of that point. Now, in the case of any other force but the gravitational, there would be no obstacle to separating out the inertial from the non-gravitational effects. For electrically charged particles, for example, one would merely vary the ratio of electric charge to inertial mass: The electric force would vary with this ratio, the inertial force would not. But the ratio of gravitational charge (= gravitational mass) to inertial mass is just what *cannot* be varied—the invariance of that ratio is the primary empirical basis of the equivalence principle.

So the answer to our question is: Once we adopt the field point of view about gravitation, there is no way (locally) to distinguish inertial from gravitational effects.

We have to recognize that there is an *inertio-gravitational field*, and that how this field divides up into inertial and gravitational terms is not absolute (i.e., frame-independent), but depends on the state of motion (in particular the acceleration) of the frame of reference being used. Indeed, we see that, by choosing the value of $\boldsymbol{A}$ to coincide numerically with the value of $\boldsymbol{g}$ at some point, we can make the total inertio-gravitational field vanish at that point. Indeed, this is why we did not call it an inertio-gravitational force: Although the values of their components with respect to some frame of reference can change depending on the state of motion of that frame, non-vanishing force fields at a point, such as the electric and magnetic fields making up the electromagnetic field, cannot be made to all vanish by any change of reference frame.

Another consequence of our new, equivalence-principle viewpoint is that a basic distinction between inertial and linearly accelerated frames of reference is no longer tenable. Any rigid non-rotating frame of reference is just as good as any other.

Let us now inventory what is left after we adopt this new viewpoint:

1. the absolute time, assumed to be measurable by ideal clocks; its measurement is unaffected by the presence of an inertio-gravitational field (compatibility of chronometry with the inertio-gravitational field);
2. Euclidean geometry, which holds within each frame in the class of three-dimensional, non-rotating frames of reference; it is assumed to be measurable with ideal measuring rods; its measurement is unaffected by the presence of the inertio-gravitational field (compatibility of geometry with the inertio-gravitational field).
3. Since gravitation and inertia are no longer (absolutely) distinguished (i.e., gravity is no longer regarded as a force), the set of "force-free" inertial motions is replaced by a set of "force-free" inertio-gravitational motions. One of these is determined by specifying a velocity vector at a point of space and an instant of time. The vector is then the tangent to the "freely falling motion" through the point at this instant.
4. While the inertio-gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$ is not absolute (i.e., it depends on the frame of reference used, and only behaves like a vector with respect to transformations within a given frame of reference), its spatial derivatives $\partial_m g^n(\boldsymbol{r}, t)$ are independent of the (non-rotating) reference frame. Physically, these differential gravitational forces are usually designated as the tidal forces, since they are responsible for the tides, among other effects. The matrix of these quantities determines the relative acceleration of two freely falling test particles, i.e., the acceleration of one particle with respect to the other. The components of the tidal forces therefore may be evaluated by measurement of the components of this relative acceleration.

6. THE NEWTONIAN CONNECTION

Now we are ready to make the transition to the four-dimensional point of view, in which a point of space-time is specified by the four coordinates (t, x^1, x^2, x^3) or $(t, \boldsymbol{r})$ for short, where x^1, x^2, x^3 are the Cartesian coordinates of the point with respect to some non-rotating frame of reference and t is the absolute time.[19] We shall refer to these as *adapted coordinates* for this frame of reference. The absolute time gives a *foliation* of space-time, i.e., a family of non-intersecting hypersurface that fills the space-time. In the adapted coordinate system the foliation consists of the hypersurfaces $t = \text{const}$. A vector is said to be *space-like* if it is tangent to a hypersurface of the foliation; a vector is *time-like* if it is not space-like. Any curve, the tangent vector to which is always time-like, is a *time-like* curve, with a similar definition for spacelike curves. In adapted coordinates a vector is time-like if it has a non-vanishing time component, space-like if it does not.

We can use any (three-)velocity field $\boldsymbol{v}(t)$ to rig the hypersurfaces of constant time: Define a time-like four-velocity field $V(t)$, with t-component = 1 and spatial components equal to those of $\boldsymbol{v}(t)$ in adapted coordinates. Thus, $V(t)$ defines a congruence of time-like curves that fills space-time. Indeed, we need merely select *one* such time-like curve $V(t)$ and then parallel propagate it along each hypersurface $t = \text{const}$ to get this congruence. In particular, the paths of the points $x^1, x^2, x^3 = \text{const}$, parametrized by the absolute time t, constitute such a congruence; Euclidean geometry holds for these spatial coordinates at all times. Thus we have specified the chronometry and the geometry of the initial frame of reference using the adapted coordinates.

Any $V(t)$ field provides a rigging of each hypersurface (see the discussion of rigged hypersurfaces in the Appendix). Just as a rigging was needed to go from the flat affine connection of the enveloping space to the non-flat affine connection of a hypersurface embedded in it, a rigging is needed here to relate the flat (Levi-Civita) connection on each Euclidean hypersurface to the four-dimensional non-flat connection that we want to define for space-time as the mathematical representation of the inertio-gravitational field.

Indeed, we can define a unique symmetric, four-dimensional affine connection on the space-time by requiring that it satisfy the following conditions:

1. The absolute time is the affine parameter for all time-like geodesic paths. A geodesic path that is time-like at any of its points is time-like at all its points.
2. There is a flat, Euclidean connection on each (three-dimensional) hypersurface of the foliation. Hence, the Euclidean distance is the affine parameter for each space-like geodesic path. A geodesic path that is space-like at any of its points is space-like at all its points.

19 We shall designate a time component by a sub- or superscript "t," and spatial components by sub- or superscript "$i, j, k\ldots$" or other lower-case Latin letters.

3. The three-dimensional and the four-dimensional treatments of the spatial geometry on each hypersurface are consonant with each other: The Euclidean (flat) three-dimensional affine connection on each hypersurface of some frame of reference coincides with the connection induced on that hypersurface by the four-dimensional connection when that hypersurface is rigged with *any* time-like $V(t)$ field.[20]
4. Parallel transport of any space-like vector is path-independent. By picking an orthonormal triad $\boldsymbol{e}_i$ of such vectors at some point on an initial hypersurface of the foliation, and parallel transporting the triad along any time-like curve with tangent vector $V(t)$, a frame of reference is generated: Once it is parallel transported to a point on another hypersurface of the foliation, the triad can be propagated to any other point of the hypersurface by (path-independent) parallel transport.
5. If we add any $V(t)$ to the triad field e_i, now interpreted as four-vectors, we get a four- dimensional frame of reference.[21] In any such frame of reference, any path that obeys the Newtonian gravitational equation of motion of a structureless test particle shall be a time-like geodesic of the four-dimensional connection parametrized by the absolute time. The spatial projection of its four-dimensional tangent vector onto any hypersurface of the foliation will coincide with the three-velocity of the test particle on that hypersurface.

As indicated earlier, we have not attempted to give a minimal list of assumptions, each of which is independent of the others; but rather, a physically intuitively plausible list. We now proceed to derive the components of the connection in some given non-rotating frame of reference, i.e., using coordinates adapted to the tetrad of basis vectors $V(t), e_i$ that characterize this frame of reference.

The equation of a geodesic in these coordinates is (see the Appendix):

$$d^2x^\kappa/d\lambda^2 + \Gamma^\kappa_{\rho\sigma}(dx^\rho/d\lambda)(dx^\sigma/d\lambda) = 0, \quad (\kappa, \rho, \sigma = t, x^1, x^2, x^3), \tag{5}$$

where λ is an affine parameter, i.e., the (four-dimensional) tangent vector to the curve $P(\lambda)$ is equal to $dP/d\lambda$; and the components of the connection are with respect to the chosen four-dimensional frame of reference. If we consider time-like geodesics, condition 1) requires that t be an affine parameter for all of them. The four-velocity $dx^\rho/d\lambda$ will thus have components $(1, \boldsymbol{v})$ in the adapted coordinate system, where $\boldsymbol{v}$ is the three-velocity of the particle. Considering only the t-component of eq. (5) for the moment, in adapted coordinates it takes the form:

---

20 If the requirement is fulfilled for one such field it is fulfilled for any such field, since two such fields can only differ by a *space-like* acceleration vector field. So the transition from one non-rotating frame of reference to another, which corresponds mathematically to a change of $V(t)$ field, does not affect the result.

21 Note that any such $V(t)$ field commutes with the three e_i fields, which commute with each other, so that they form a holonomic basis; so coordinates adapted to this basis will always exist.

$$d^2t/dt^2 + \Gamma^t_{tt} + 2(\Gamma^t{}_{ti})v^i + (\Gamma^t_{ij})v^iv^j = 0, \tag{5a}$$

and since the first term vanishes, the only way that eq. (5a) can hold for all values of v^i is if Γ^t_{tt}, Γ^t_{ti} and Γ^t_{ij} all vanish in the adapted coordinate system. In other words, these are the mathematical conditions that assure the compatibility of the chronometry and the inertio-gravitational field. Physically, this means that an ideal clock moving around in the inertio-gravitational field will always measure the absolute time.

Conditions 2, 3, and 4 now demand that the three space-like vectors e_i, which lie along the coordinate axes and thus have components δ^μ_i in adapted coordinates, have vanishing covariant derivates with respect to both the Euclidean (flat) three-dimensional connection on each hypersurface, and the non-flat inertio-gravitational four-dimensional connection. By a similar argument to that above, these conditions result in the vanishing of Γ^m_{ti} and Γ^m_{ni} in the adapted coordinate system. In other words, these are mathematical conditions that assure the compatibility of the geometry and the inertio-gravitational field. Physically, this means that an ideal measuring rod moving around in the inertio-gravitational field will always measure the Euclidean distance.

Condition 5 now fixes the values of the only remaining non-vanishing components of the affine connection, Γ^m_{tt}, in the adapted coordinate system. Returning to eq. (5), its spatial components in the adapted coordinate system now take the form:

$$d^2x^m/dt^2 + \Gamma^m_{tt} = 0, \tag{5b}$$

all other terms in the equation vanishing because of the previously-established vanishing of the other components of the connection. We see that we need merely set:

$$\Gamma^m_{tt} = -g^m(\boldsymbol{r}, t) \tag{6}$$

in the adapted coordinates in order to have the geodesic equation coincide with the equation of motion of a particle in the gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$.

We have now fixed all the components of the symmetric affine connection in the adapted coordinate system. We need merely apply the general transformation law for the components of the connection under a coordinate transformation $x^{\kappa'} = x^{\kappa'}(x^\kappa)$:

$$\Gamma'^{\kappa}_{\mu\nu} = \Gamma^\kappa_{\mu\nu}(\partial x^{\kappa'}/\partial x^\kappa)(\partial x^\mu/\partial x^{\mu'})(\partial x^\nu/\partial x^{\nu'}) + \partial^2 x^{\kappa'}/\partial x^\mu \partial x^\nu, \tag{7}$$

to the equations for a linearly accelerated transformation (see eq. (2) of Section 5):

$$x^{m'} = x^m + R^m(t), \tag{8}$$

in order to see that the components of the connection transform correctly; i.e, that all the components but Γ'^m_{tt} continue to vanish, and the Γ^m_{tt} transform just like the components of $\boldsymbol{g}$ under such a transformation (see Section 5, eq. (4)). If we carry out a transformation to a rotating system of coordinates, the transformation of the components of the connection introduces terms that correspond to the Coriolis and centripetal "inertial forces" that must be introduced in a rotating coordinate system. To get the form of the components of the connection in an arbitrary coordinate system, one need merely apply eq. (7) to an arbitrary coordinate transformation.

What about the tidal forces, which as mentioned above are absolute? They are represented by the appropriate components of the Riemann tensor, which can be computed from the Newtonian inertio-gravitational connection. Since they are components of a tensor, they indeed possess an absolute character, in the sense that if the components do not all vanish at a point, no change in frame of reference at that point can make them all vanish. These components of the Riemann tensor enter into the equation of geodesic deviation, which describes in four-dimensional tensorial form the relative acceleration of two particles falling freely in the inertio-gravitational field; but I shall not enter into details here.

Rather, I turn to the question of the field equations for the inertio-gravitational field. The Newtonian field $\boldsymbol{g}$ obeys the field equation:

$$\nabla \cdot \boldsymbol{g} = 4\pi G\rho, \tag{9}$$

where G is the Newtonian gravitational constant, ρ is the mass density of the material sources of the gravitational field, and $\nabla \cdot \boldsymbol{g}$ is the trace of the tidal force matrix. If one works out the components of $R_{\mu\nu}$, the contracted Riemann or Ricci tensor, in the adapted coordinates, it turns out that only R_{tt} is non vanishing, and it equals $-\nabla \cdot \boldsymbol{g}$. So $R_{tt} = -4\pi G\rho$ and all other components =0 in the adapted coordinates. The only remaining problem is to write this result as a tensorial equation, independent of coordinate system; but this is easily solved by introducing a covariant vector field T_μ, such that in adapted coordinates $T_\mu = \partial_\mu t = \delta^\mu_t$. The gravitational field equations now take the tensorial form:

$$R_{\mu\nu} = 4\pi G\rho T_\mu T_\nu, \tag{10}$$

which is clearly of the same form in all coordinate systems.

In a more complete treatment,[22] one would have to go a step further: the Newtonian gravitational field $\boldsymbol{g}$ can be derived from a gravitational potential function ϕ: $\boldsymbol{g} = -\nabla\phi$, and this condition can be expressed intrinsically in terms of the properties of the corresponding Riemann tensor (the tidal force matrix introduced in Section 5, which is closely related to certain components of the Riemann tensor, becomes symmetric). Now ϕ plays an important role in taking the Newtonian limit of general relativity, but since we shall not discuss this issue, I can forego entering into further consideration of details.

The non-dynamical Newtonian chrono-geometrical structures, consisting of the absolute time and the relative spaces of the family of non-rotating frames of reference, are unmodified by the presence of gravitation. Mathematically, they are represented by a closed temporal one-form (the T_μ introduced above) and a trivector field

22 See (Stachel 2003b) for such a treatment.

whose transvection with the one-form vanishes (the e_i introduced above), from which a degenerate (rank 3) spatial "metric" may be constructed ($=\delta^{ij}e_ie_j$).

However, the compatible flat inertial structure of Newton's theory is modified. It becomes a dynamical structure, the Newtonian inertio-gravitational field, which remains compatible with the chrono-geometrical structures. Mathematically, it is represented by a symmetric affine connection (the Newtonian connection $\Gamma^{\kappa}_{\rho\sigma}$ discussed above), which can be derived from a "connection potential" (the ϕ discussed above). Its contracted Riemann tensor obeys field equations that relate it to the masses acting as its source (eq. (10) above). The compatibility of this connection with the chrono-geometrical structure means, as noted earlier, that clocks and measuring rods freely falling in the inertio-gravitational field still measure absolute temporal and spatial intervals, respectively. Mathematically, this is expressed by the vanishing of the covariant derivatives of the temporal one-form and degenerate spatial "metric" with respect to the Newtonian connection.

7. SOME MYTHICAL HISTORY: NEWSTEIN MEETS WEYLMANN

Once the concept of affine connection has been developed and the Riemann tensor geometrically interpreted in terms of parallel transport around closed curves, this version of Newton's theory—which converts gravitation from a force that pulls bodies off their (non-dynamical) inertial paths, into a (dynamical) modification of the (inertial) affine connection—is almost immediately suggested by the equality of gravitational and inertial mass. Indeed, shortly after the mythical mathematician Weylmann formulated the concept of affine parallelism, his equally mythical physicist colleague Newstein developed this reinterpretation of Newtonian gravitational theory. Brooding on the equality of gravitational and inertial mass, he became convinced of the essential unity of gravitation and inertia. Originally, he expressed this insight in the usual three-plus-one language of physics, treating space and time separately (see Section 5). He considered uniformly accelerated frames of reference in the absence of gravitation (the Newstein elevator!), and decided it was impossible to distinguish such a frame of reference from a non-accelerated frame with a constant gravitational field. This led him to consider transformations between linearly accelerated frames of reference.

He was puzzled by the strange transformation law that he had to introduce for the gravitational "force," which no longer behaves like a vector under such transformations. At some point he turned to Weylmann, who soon realized that the gravitational "force" transforms like the Γ^{m}_{tt} components of a four-dimensional affine connection, and that Poisson's law for the gravitational potential could be written as an equation linking the Ricci tensor of the connection with its material sources (see Section 6). In the now-famous Newstein-Weylmann paper, the two developed a four-dimensional geometrized formulation of Newtonian gravitation theory, which generalized Newtonian chrono-geometry to include linearly accelerated frames and a dynamized inertio-gravitational connection field, but still included the concept of absolute time.

In so far as they took any notice of this work, their contemporaries regarded it as an ingenious mathematical *tour-de-force*. But, since it had no new physical consequences, it did not much impress Newstein's positivistically-inclined physics colleagues.

Weylmann analyzed the invariance group of the new theory, which is much wider than that of the older Newtonian kinematics. The privileged role of the inertial frames of reference in Newton's theory, just beginning to be realized thanks to the work of Lange and Neumann, was lost in the new interpretation of gravitation.While rotation remained absolute (in the sense that all components of the connection representing centrifugal and Coriolis forces could be made to vanish globally by a coordinate transformation), all linearly accelerated frames of reference were now equal, and the significance of this occasioned a discussion among a few philosophers of science who concerned themselves with the foundations of mechanics. Ernst Mach added a few lines about Newstein to the latest edition of his *Mechanik*.

8. MORE MYTH: EINSTEIN CONFRONTS NEWSTEIN

Perhaps this is where Albert Einstein first read about Newstein's work. At any rate, in 1907, pursuant to his commission to write a review article on the physical consequences of his 1905 work on the relativity principle (now becoming known as the theory of relativity),[23] he turned his attention to gravitation, and (like Newstein) was struck by the equality of gravitational and inertial mass. He realized that, as a consequence, in Newtonian mechanics there is a complete equivalence between an accelerated frame of reference without a gravitational field and a non-accelerated frame of reference, in which there is a constant gravitational field. He soon generalized this to what he later called the principle of equivalence: There is no physical difference (mechanical or otherwise) between the two frames of reference.[24]

Recalling what he had read about Newstein, Einstein realized that he had rediscovered the loss of the privileged role of inertial frames once gravitation is taken into account. Like Newstein, he became convinced that inertia-cum-gravitation must be represented mathematically by an affine connection; but now this representation somehow must be made compatible with the new chronogeometry he had developed in his 1905 theory.[25] He first tried to preserve the non-dynamical nature of this chrono-geometrical structure—which Minkowski soon expressed in terms of a four-dimensional pseudo-Euclidean geometry—by developing various special-relativistic gravitational theories that incorporated the unity of gravitation and inertia by the very fact that they were based upon an affine connection. But the Riemann tensor of the inertio-gravitational connection in each of these theories was non-vanishing, while

---

23 For a translation of this paper, see (Stachel 1998).

24 Aside from the first sentence, this paragraph is a summary of the actual historical circumstances of Einstein's first work on gravitation, see (Einstein 1907). The fantasy begins in the next paragraph.

25 In the frame bundle language, the physically preferred subgroup of the general linear group had to be changed from the Newtonian group to the Lorentz group.

the metric-affine structure of Minkowski space-time is flat. Physically, this meant that the inertio-gravitational and chrono-geometrical structures were not compatible: Good clocks and measuring rods, as defined by the chrono-geometrical structure, did not keep the proper time or measure the proper length when moved about in the gravitational field.

While this could be "explained away" as due to a universal distorting effect of gravitation on all measuring rods and clocks, something about such an explanation disturbed him. Since the effect was universal, the "true" Minkowski chronogeometry could be shown to have no physically observable consequences.

Finally, he realized what was bothering him: This type of explanation was all too similar to Lorentz's interpretation of the Lorentz transformations: Galilean chronogeometry is the "true" one; but the universal effect of motion through the absolute (aether) frame of reference exerts a universal effect on all physical processes that prevents any physically observable consequences of this motion. What was the way out of this new unobservability dilemma?

Suddenly the answer struck him: If he required compatibility between the inertio-gravitational and chrono-geometrical structures, the problem would disappear, just as it had in Newstein's reinterpretation of Galilean kinematics. Good measuring rods and clocks, as defined by such a chrono-geometrical structure, would measure the true proper lengths and times wherever they were placed in the inertio-gravitational field. But there was a price to pay for this compatibility: The chrono-geometrical field could no longer be flat. It would have a curvature attached to it in the Gaussian sense, the one that Riemann originally had generalized from two to an arbitrary number of dimensions. In this theory, the Riemann tensor would have two distinct (but compatible) interpretations: as the curvature of a connection, associated with parallel transport and the equation of geodesic deviation; and as the curvature of a pseudo-metric, associated with the Gaussian curvature of each of the two-dimensional sections at any point of space-time.

And of course, since metric and connection were now compatible, this implied that the components of the connection with respect to any basis were numerically equal to the Christoffel symbols of the metric with respect to that basis. And since the connection is a dynamical field, the metric would also have to become a dynamical field. In contrast to the Newsteinian case, where the chrono-geometry remained non-dynamical, in the Einsteinian case, there are no non-dynamical space-time structures. The bare manifold remained absolute in a certain sense;[26] but then, it had no physical characteristics other than dimensionality and local topology unless and until the iner-

26 I say this because, in actual fact, the global topology of the manifold is not given before the metric-cum-connection field, as implied in so many presentations of general relativity. One actually solves the Einstein field equations on a small patch, and then looks for the maximal extension of that patch compatible with the given metric. Certain criteria for compatibility must be given before the question of maximal extension(s) becomes meaningful, of course. For discussion of this topic, see (Stachel 1986; 1987).

tio-gravitational cum chronogeometrical field was impressed upon it. Least of all do the points of the manifold represent physical events before imposition of a metric.[27]

The new, dynamical theory of space-time structures had a number of novel physical consequences, and Einstein soon became world-famous—but you know the rest of the story.

9. SOME REAL HISTORY: EINSTEIN WITHOUT NEWSTEIN

Unfortunately, the last section was a historical fable, and the real Einstein had to work out the general theory of relativity in the absence of the concept of affine connection—an absence which, as suggested in Section 2, played a fateful role in the actual development and subsequent history of the theory. It took Einstein without Newstein seven years to develop the general theory of relativity *after* he had adopted the equivalence principle as the key to a relativistic theory of gravitation. Rather than tell the entire story of the many genial steps and equally numerous missteps on Einstein's road from special to general relativity,[28] I shall here just highlight some of the most fateful consequences of the absence of the connection.

First of all, it is important to realize that the tensor calculus, as originally developed by Christoffel, Ricci, Levi Civita and others, was a branch of invariant theory, with only tenuous ties to geometry.[29] Einstein's introduction of the metric tensor field as the mathematical representation of both the chrono-geometry of space-time and the potentials for the gravitational field did not carry with it most of the geometrical implications that we take for granted today. Insofar as it did carry geometrical implications, notably in fixing the geodesics of the manifold, this had to do with the interpretation of geodesics as the shortest paths (or rather longest, for time-like paths—the twin paradox) in space-time. The interpretation of geodesics as the straightest paths in space-time, more important for the understanding of the gravitational field—in particular, the interpretation of the Riemann tensor in terms of the equation of geodesic deviation—had to await the work of Levi Civita and Weyl on parallelism discussed in Section 3.[30] Curvature, in other words, was given the Gauss-Riemann interpretation, rather than the interpretation as the tendency of geodesics to coverge (or diverge), leading to its association with tidal forces.

---

27 For discussion of the hole argument, which bears on this point, see (Stachel 1993) and references therein.

28 See the first two volumes of this series on the development of general relativity. For earlier accounts by this author and others, see (Stachel 1995) and the references therein.

29 "The calculus developed by Gregorio Ricci in the years 1884–1887 had its roots in the theory of invariants, therefore it naturally lacked a geometrical outlook or interpretation, and was so intended by Ricci" (Reich 1992, 79). For the history of the tensor calculus, see (Reich 1994).

30 Interestingly, this interpretation was anticipated by Hertz in his geometrical version of mechanics. See (Hertz 1894) and, for a discussion of the 19th century tradition of geometrical interpretations of mechanics, (Lützen 1995a; 1995b).

It is often said that Einstein, with the help of Grossmann, found ready-to-hand the mathematical tools he needed to develop general relativity: Riemannian geometry and the tensor calculus. But this statement must be taken with a large grain of salt. It would be more correct to say that he had to make do with the tools at hand, with important negative consequences for the development of the theory, and—more importantly for us now—with negative consequences for the interpretation of the theory that continue to exert their effects to this day.[31]

To give two concrete examples of this negative influence on Einstein's work:

1. Until late in 1915, he regarded the derivatives of the metric tensor, rather than the Christoffel symbols, as the mathematical representative of the gravitational-cum-inertial field.[32] In Einstein 1915, he finally corrected this error:

   > These conservation laws [the vanishing of the covariant derivative of the stress-energy tensor] previously misled me into regarding the quantities $1/2\sum_{\mu} g^{\tau\mu}\partial g_{\mu\nu}/\partial x_{\sigma}$ as the natural expression for the components of the gravitational field, although in the light of the formulas of the absolute differential calculus it seems more obvious to introduce the Christoffel symbols instead of these quantities. This was a fateful prejudice (Einstein 1915, 782).

   The reason why this error was so fateful is that it mislead Einstein in his search for the gravitational field equations, a search that took over two years *after* he had adopted the metric tensor field as the mathematical representation of gravity.[33]

2. From 1912 onwards, Einstein expected that, in the Newtonian limit of general relativity, the spatial part of the metric field tensor would remain flat and that the g_{oo} component of the metric would reduce to the Newtonian gravitational potential. Correctly understood, in terms of a formulation of the theory taking the Newtonian limit of both the connection and the metric, these expectations are fulfilled. But one cannot properly take the Newtonian limit of general relativity without the concept of an affine connection, and the corresponding affine reformulation of Newtonian theory discussed in Section 6. Indeed, the problem of correctly taking the Newtonian limit of general relativity only began to be solved in (Friedrichs 1927), and the process was not completed in all details until (Ehlers 1981). In the absence of the affine approach, more-or-less heuristic detours through the weak-field, fast motion (i.e., special-relativistic) limit followed by a slow motion approximation basically out of step with the fast-motion approach, had to be used to "obtain" the desired Newtonian results.[34]

---

31 Perhaps the first such negative influence on work done after the final formulation of the general theory is the ultimate failure of Lorentz's attempt to give a coordinate-free geometrical interpretation of the theory. I thank Dr. Michel Janssen for pointing this out to me. For an account of Lorentz's attempt, see (Janssen 1992).

32 See (Einstein and Grossmann 1913, 7), and (Einstein 1914, 1058), for examples.

33 For details see volume 1 of this series on the development of general relativity.

34 See (Stachel 2003b) for more details.

Einstein originally thought that he knew the form of the weak field metric in the static case. It involved a spatially flat metric tensor field, with only the g_{oo} component of the metric depending on the coordinates. He used this form of the static metric as a criterion for choosing the gravitational field equations: This form of the metric had to satisfy the field equations, which led to a disastrous result: No field equation based on the Ricci tensor had this form of the static metric as a solution, and Einstein abandoned the Ricci tensor for over two years![35] Had he known about the connection representation of the inertio-gravitational field, he would have been able to see that the spatial metric can go to a flat Newtonian limit, while the Newtonian connection remains non-flat without violating the compatibility conditions between metric and connection. As it was, using the makeshift technique described above to get the Newtonian result, he was amazed to find that the spatial metric is non-flat. Even today, almost all treatments of the Newtonian limit of general relativity are still based on this makeshift approach that employs only the metric tensor.

10. CONCLUSION

The moral of this story is that general relativity is primarily a theory of an affine connection on a four-dimensional manifold, which represents the inertio-gravitational field. The other important space-time structure is the metric field that represents the chrono-geometry; and the peculiarity of general relativity is that the compatibility conditions between metric and connection—or in physical terms, between inertio-gravitational field and chrono-geometry—uniquely determine the connection in terms of the metric. In teaching the subject, emphasis should be put on the connection from the beginning. This can be done easily by presenting the affine version of Newtonian gravitation theory before discussing general relativity. But most textbooks still start from the metric and introduce the connection later via the Christoffel symbols in a way that does not stress the basic role of the connection.[36] Now that gauge fields have come to dominate quantum field theory, it is more important than ever to emphasize from the beginning how general relativity resembles these Yang-Mills type theories, as well as how it differs.[37]

---

35 For details, see (Stachel 1989; Norton 1984) and volume 1 of this series.

36 It is indicative of current interests that (Darling 1994), the only elementary mathematical textbook I know that introduces the connection first, does not even mention the application to gravitation theory, but concludes with a chapter on "Applications to Gauge Field Theory" (pp. 223–250).

37 The basic difference is that the affine connection lives in the frame bundle (see Section h of the Appendix), which is soldered to the space-time manifold. The symmetries of the fibres are thus induced by space-time diffeomorphisms. On the other hand, the Yang-Mills connections live in fibre bundles, the fibres of which have symmetry groups that are independent of the space-time symmetries (internal symmetries). For further discussion, see (Stachel 2005).

ACKNOWLEDGEMENTS AND A CRITICAL COMMENT

I thank Dr. Jürgen Renn for a thoughtful reading of this paper, and many helpful suggestions for its improvement. I thank Dr. Erhard Scholz for his careful critique of the paper. While agreeing with its basic viewpoint, he made some critical comments on my treatment of Grassmann and the mythical Weylmann. With his kind permission I quote them:

> The (historical) *lineale Ausdehnungslehre* was so much oriented towards the investigations of linear geometric structures and their algebraic generalization that there was a deep conceptual gulf between Grassmann's approach and Riemann's differential geometry of manifolds, which could only be bridged after a tremendous amount of deep and hard work. I do not see in Grassmann's late attempt to understand the algebraic geometry of curves and surfaces in terms of his *Ausdehnungslehre* a step that might have led him even somewhat near to a generalization of parallel transport in the sense of differential geometry. In "real history" there was no natural candidate for "Weylmann."
>
> ..... So, in short, your Newstein paper is an interesting thought experiment discussing the question of what would have happened if history had gone other than it did. In doing so, and following your line of investigation, we might find more precise answers as to why there was, e.g., still a long way to go from Grassmann to a potential "Weylmann." This is contrary to your intentions, I fear, but I cannot help reading your paper that way.

Rather than going contrary to my intentions, his remarks raise a most important question that supplements my approach to alternate histories: Given that we can invent various alternatives to the actual course of events, can one attach a sort of intrinsic probability to these various alternatives? I mean probability in the sense of a qualitative ranking of the probability of the alternatives rather than attaching a numerical value to the probability of each. In a truly "postmature" case, the ranking of the actual course of events would be lower than that of at least one of the alternatives. For example, the probability of a direct mathematical route from Riemann's local metric to Levi-Civita's local metrical parallelism would rank higher than the probability of the actual route via physics through Einstein's development of general relativity. Dr. Scholz makes a strong case for ranking the probability of the actual course of events from Grassmann's affine spaces to Weyl's affine connection higher than the probability of the step from Grassmann to Weylmann in my myth. I shall not pursue this issue further here, but again thank Dr. Scholz for comments that raise it in the context of my paper.

APPENDIX: RIEMANNIAN PARALLELISM AND AFFINELY CONNECTED SPACES

I shall review the concepts of parallelism in Euclidean and affine spaces, and their generalization to non-flat Riemannian and affinely connected spaces, respectively. I shall emphasize material needed to understand the historical and mathematical discussion in the Sections 3–6 and Newstein's mythical history in Section 7. Those familiar with the mathematical concepts may refer to the Appendix as needed when reading Sections 4–7.[38]

a. affine and Euclidean spaces. The familiar concept of parallelism in Euclidean space can easily be extended from lines to vectors: two vectors at different points in that space are parallel if they are tangent to parallel lines. We say that two Euclidean vectors are equal if they are parallel and have the same length as defined by the metric of Euclidean space. But, as we shall soon see, the concepts of parallelism and equality of parallel vectors retain their significance when we abstract from the metric properties of Euclidean space to get an affine space.

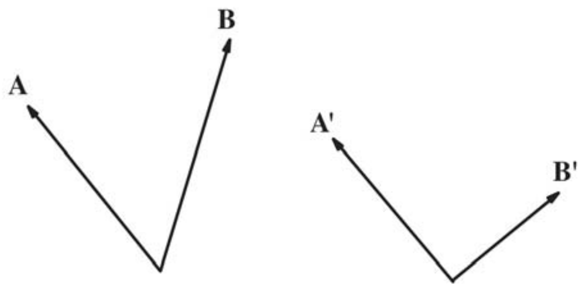

Figure 1: Any pair of non-parallel vectors $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *can be transformed into any other pair* $\boldsymbol{A}'$ *and* $\boldsymbol{B}'$ *by an (active) affine transformation.*

The properties of Euclidean geometry may be defined as those that remain invariant under transformations of the Euclidean group, consisting of translations $T(3, R)$,[39] and of rotations $O(3, R)$ about any point in space.[40] A translation is a point transformation that takes the point P into the point $P + \boldsymbol{v}$, where $\boldsymbol{v}$ is any vector. A rotation is a point transformation with a fixed point P that takes the point

38 However, in contrast to more familiar treatments, I shall define connections in terms of frame bundles, a concept that I shall introduce informally, following (Crampin and Pirani 1986, chaps. 13–15), which may be consulted for more details.

39 I shall use the notation (n, R) to denote a group acting on a real n- dimensional space.

40 I shall give the active interpretation of all geometrical transformations: The transformations act on the points of the space in question, taking each point into another one. The idea of defining a geometry by the group of transformations that leave invariant all geometric relations goes back to (Klein 1872).

$P + \boldsymbol{r}$ into the point $P + \mathcal{O}r$ where $\mathcal{O} \in O(3, R)$ is an orthogonal transformation. The translations are clearly metric-independent; but the orthogonal transformations, being the linear transformations that preserve the distance between any pair of points, clearly do depend on the metric.

If we relax the condition that a linear transformation $\mathcal{L}$ preserve distances, and merely demand that it have a non-vanishing determinant), then $\mathcal{L} \in GL(3, R)$, the group of general linear or affine transformations. Together with the translations, they form the affine group that defines an affine geometry.[41] Parallelism of lines and vectors and the ratio of the lengths of parallel vectors (and hence the equality of two such vectors) being invariant under the affine group, are meaningful affine concepts. The (Euclidean) length of any vector is changed by an affine transformation with non-unit determinant, so it is not a meaningful affine concept.

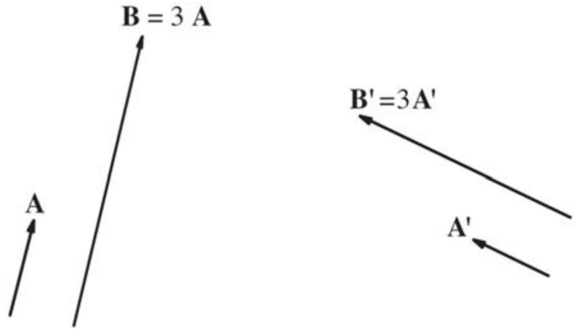

Figure 2: Any pair of parallel vectors **A** *and* **B** *can be transformed into any other pair of parallel vectors* **A'** *and* **B'** *with the same ratio by an (active) affine transformation.*

In order to determine the action of an affine transformation $\mathcal{L}$ on any vector $\boldsymbol{v}$ at some point of an n- dimensional affine space, we need merely define its action on a basis or linear frame $\boldsymbol{e}_i$ at that point, consisting of n linearly-independent vectors:

$$\boldsymbol{e}_j{}' = \mathcal{L}^i_j \boldsymbol{e}_i, \tag{11}$$

where $\boldsymbol{e}_j{}'$ is the new basis produced by the action of $\mathcal{L}$ on $\boldsymbol{e}_i$, and $\mathcal{L}^i_j$ is the matrix representing the action of $\mathcal{L}$ on some basis. (Here and throughout, we have adopted the summation convention for repeated indices, which range over the appropriate number of dimensions—here 1 ,..., n.)

If we want to restrict ourselves to Euclidean geometry and the orthogonal group, we may restrict ourselves to orthonormal bases or frames:

41 For a discussion of affine and metric spaces, with a view to the generalizations needed below, see (Crampin and Pirani 1986, chaps. 1 and 7). For these generalizations, see chaps. 9 and 11.

$$e_i \cdot e_j = \delta_{ij}, \tag{12}$$

where the dot symbolizes the Euclidean scalar product of two vectors, and to orthogonal changes of bases:

$$e_j{}' = \mathcal{O}^i_j e_i, \qquad e_i{}' \cdot e_j{}' = \delta_{ij}. \tag{13}$$

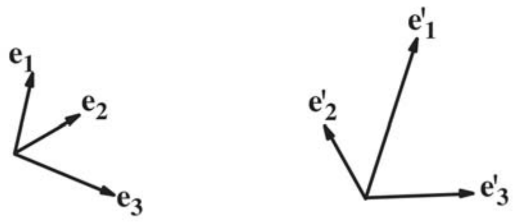

Figure 3: A (homogenous) affine transformation is defined by its action on a basis (or linear frame) e_A *of the affine space.*

Once we have chosen a basis at one point of an affine (or Euclidean) space, we can take as the basis at any other point of space the set of basis vectors equal and parallel to the original basis, thereby setting up a field of bases or *linear frames* over the entire space.

b. frame bundles. On the other hand, we can consider the set of all possible bases or linear frames at a given point of space. As is clear from eq. (11), in an affine space these frames are related to each other by the transformations of $GL(n, R)$. The set of all frames, together with the structure that the n-dimensional affine group imposes on them, is said to form a *fibre* over the point in question. Similarly, in Euclidean space, the set of all possible orthonormal frames at a point has a structure imposed on it by $O(n, R)$, the n-dimensional orthogonal group (see eq. (13)).

The set of all possible frames at every point of a space together with the space itself form a manifold that is called the *bundle of linear frames* or, more simply, the *frame bundle*. This is a special case of the more general concept of a *fibre bundle*.[42] The original space, which is affine or Euclidean in our examples but capable of generalization to any manifold, is called the *base space* of the fibre bundle; each *fibre* also need not be composed of linear frames, but may have a more general structure (below we shall consider fibres composed of tangent spaces). But there is always a *projection operation* that takes us from any fibre of the bundle to the point of the base

42 For fibre bundles in general and the frame bundle in particular, see (Crampin and Pirani 1986, chaps. 13 and 14).

space at which the fibre is located. A fibre bundle is called *trivial* if it is equivalent to the Cartesian product of a base manifold times a single fibre with a structure on it. The frame bundles we have been considering are trivial, since they are equivalent to the product of an affine (or Euclidean) space times a frame fibre with the structure imposed on it by the affine (or orthogonal) group.

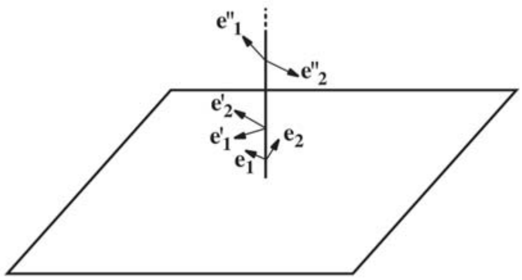

Figure 4: The set of all possible frames $\boldsymbol{e}_A$, $\boldsymbol{e}'_A$, $\boldsymbol{e}''_A$, … *at a point of the space forms a "fibre" over the point.*

A *cross-section* of the frame bundle is a specification of a particular frame on each fibre of the bundle, i.e., at each point of the base space (see Fig. 6). (The frames must vary in a smooth way as we pass from point to point, but we shall not bother here with such mathematical details.) In an affine (or Euclidean) space, the specification of a linear (or orthonormal) frame on one fibre allows us to pick out a unique parallel cross section of the entire bundle. (The last sentence just repeats, in the language of fibre bundles, something said earlier.) A change of frame on one fibre produces a change of the entire parallel cross section that is induced by an affine (orthogonal) transformation on the original fibre.

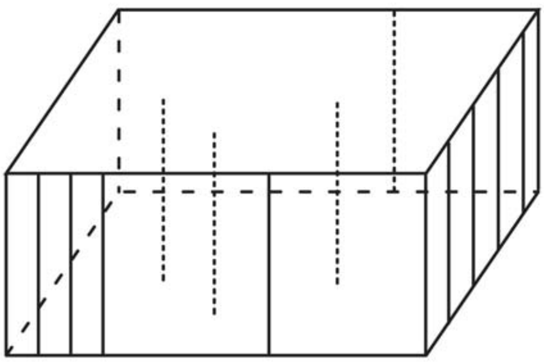

Figure 5: The fibres of a fibre bundle.

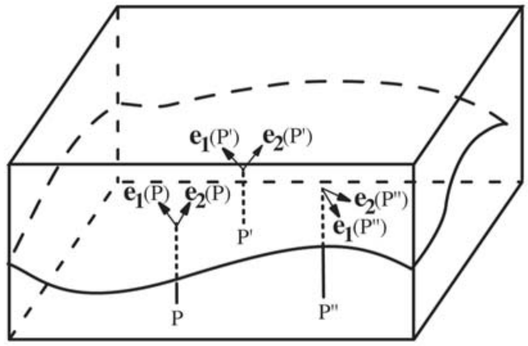

Figure 6: A "cross-section" of a frame bundle is a choice of a particular frame on each fibre of the bundle.

c. parallelism in non-flat Riemannian spaces. Now consider three-dimensional Euclidean space and some two-dimensional (generally curved) surface S in it. All vectors that are tangent to S at one of its points P form a vector space $T(P)$, called the tangent space to S at P. The collection of all such tangent spaces for all points $P \in S$ form a fibre bundle $T(S)$, called the *tangent bundle*. All vectors in $T(S)$ are intrinsically related to S,[43] and we want to define the concept of parallelism for such vectors in such a way that it will also be intrinsic to S. We cannot simply take the vector at another point P' of S that is parallel to a vector of $T(P)$ in the three-dimensional Euclidean sense: in general, that vector will not even be in $T(P')$, see Fig. 7).

We can get an idea of how to proceed by considering the case when S is a plane. The concept of parallel vectors at different points of the plane is clearly intrinsic to the plane. Consequently, the tangent spaces at each point of the plane can be identified with each other in a natural way, as can pairs of orthonormal vectors $\boldsymbol{e}_A(A = 1, 2)$ that form a basis at each point of the plane considered as a two-dimensional Euclidean space. Taken together with the unit normal vector $\boldsymbol{n}$ to the plane, the $\boldsymbol{e}_A$ form a basis for the tangent space of the three-dimensional Euclidean space.

43 These vectors can, for example, be defined as the tangent vectors to curves $C = P(s)$ lying entirely in S. We follow the usual terminology in distinguishing *curves* from *paths*, which are curves without a parametrization s.

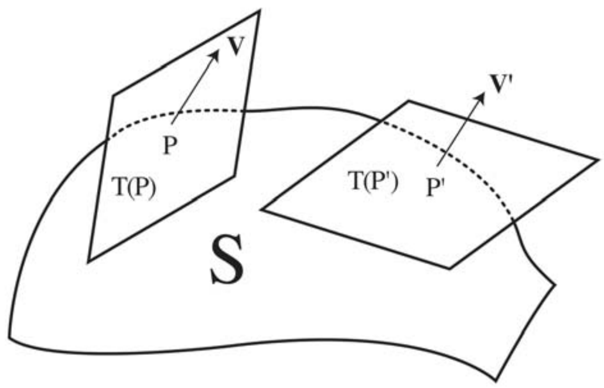

Figure 7: To define an intrinsic notion of parallelism within a surface S, we cannot use vectors that are parallel to each other in the three-dimensional sense. While V lies in the tangent plane at P, the three-dimensionally parallel vector V' does not even lie in the tangent plane at P'.

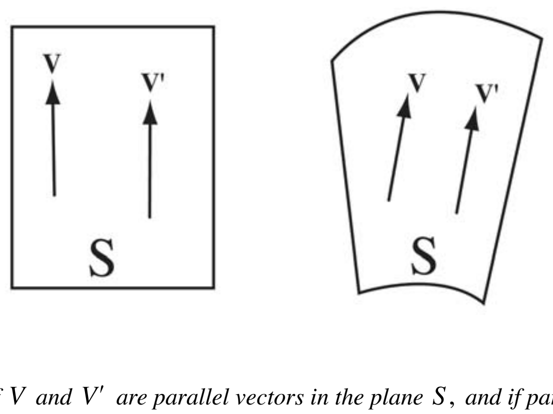

Figure 8: If V and V' are parallel vectors in the plane S, and if parallelism is intrinsic to S, then they remain parallel even when S is bent (without distortion).

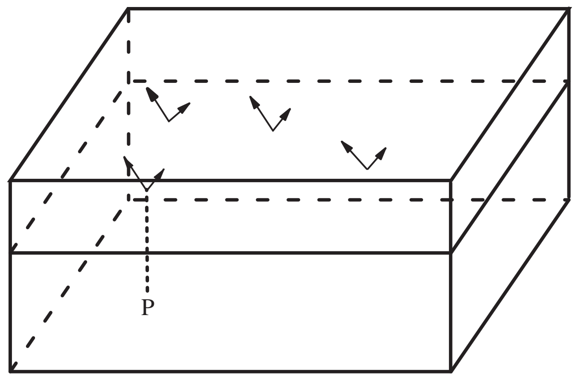

Figure 9: In an affine space, choice of a frame on one fibre picks out a unique parallel cross-section.

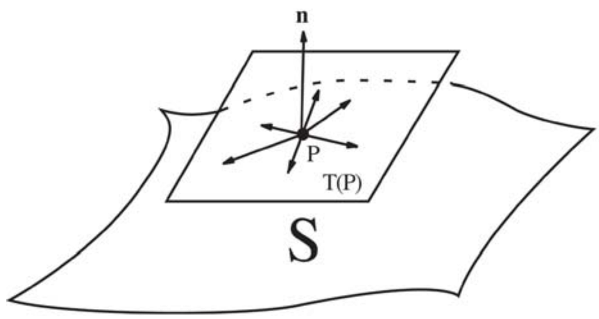

Figure 10: The tangent space $T(P)$ to a surface S at point P of the surface in Euclidean space is composed of all vectors tangent to the surface at that point. The unit normal to the tangent plane is designated by $\boldsymbol{n}$.

Now suppose we bend the plane without distorting its metric properties (i.e., the metrical relations between its points as measured *on the surface*), resulting in what is called a *developable surface*.[44] If we want the concepts of parallelism and straight line to be intrinsic to a such a surface, they must remain the same for any surface developed from the plane as they were for the plane itself. Thus, the basis vectors $\boldsymbol{e}_A$

44 Such a process of bending leaves the *intrinsic* geometry of the surface unchanged, but changes its *extrinsic* geometry. The intrinsic properties of any surface are those that remain unchanged by all such bendings; its extrinsic properties are precisely those that depend on how the surface is embedded in the enveloping Euclidean space.

at different points of the surface must still be considered parallel to each other from the intrinsic, surface viewpoint, even though they are not from the three-dimensional Euclidean point of view. Consider two neighboring points on the surface P and $P' = P + d\boldsymbol{r}$. In order to get from the tangent plane $T(P)$ at P to the tangent plane $T(P')$ at P' one must rotate the former through the angle $d\theta$ that takes $\boldsymbol{n}$ into $\boldsymbol{n}'$.[45] Thus, there must be an orthogonal transformation $\mathcal{O}$, differing from the identity $\mathcal{I}$ only by an amount that depends on P, P', $\boldsymbol{n}$ and $\boldsymbol{n}'$, or equivalently on P, $\boldsymbol{n}$, $\boldsymbol{n}'$ and $d\boldsymbol{r}$:

$$\mathcal{O} = \mathcal{I} + d\mathcal{O}, \quad d\mathcal{O} = d\mathcal{O}(P, \boldsymbol{n}, \boldsymbol{n}', d\boldsymbol{r}) \tag{14}$$

and depends linearly on $d\boldsymbol{r}$.

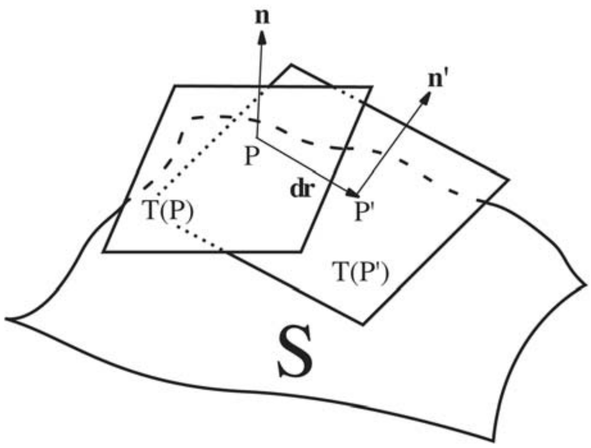

Figure 11: In order to get from the tangent plane $T(P)$ at P to a neighboring tangent plane $T(P')$ at $P' = P + d\boldsymbol{r}$, we must carry out an orthogonal transformation $\mathcal{O} = \mathcal{I} + d\mathcal{O}$ that depends on P, $d\boldsymbol{r}$, $\boldsymbol{n}$ and $\boldsymbol{n}'$.

Due to the linearity of vector spaces, the effect of this orthogonal transformation on any vector in the tangent plane to the surface can be computed once its effect on a set of basis vectors $\boldsymbol{e}_A$ in the tangent plane is known.[46] The change in each basis vector is given by:

$$\delta\boldsymbol{e}_B(P') = (d\boldsymbol{O})^A{}_B\, \boldsymbol{e}_A, \tag{15}$$

---

45 The concept of parallelism in the Euclidean space allows us to draw the vector at P that is equal and parallel to $\boldsymbol{n}'$ at P', and so define the angle $d\theta$ between $\boldsymbol{n}$ and $\boldsymbol{n}'$.

46 Note that we need the normals $\boldsymbol{n}$ and $\boldsymbol{n}'$ to define the orthogonal transformation between parallel vectors lying in the tangent planes at P and P'; but since we are only interested in the change in vectors lying in the surface we may omit $\boldsymbol{n}$ from explicit mention in eq. (5), since it is determined by the $\boldsymbol{e}_A$ and the orthonormality conditions.

where $(d\mathbf{0})^A_B\ (P, d\boldsymbol{r})$ are the elements of a matrix that determines the effect of the infinitesimal rotation on the orthonormal basis vectors.

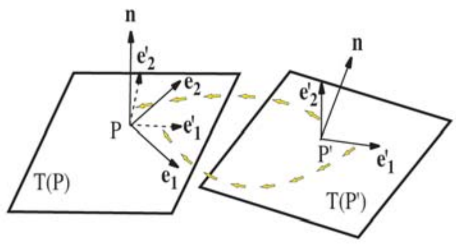

Figure 12: The effect of $d\boldsymbol{O}$ on any vector in $T(P)$ is determined by its effect on a set of basis vectors $\boldsymbol{e}_A$ of the space.

It is this connection between parallel vectors in neighboring tangent planes, given by eqs. (14) and (15), that we shall preserve for all surfaces, in particular for those that are not intrinsically plane. Since it was introduced by Levi-Civita (see Section 4), it is often called the Levi-Civita connection. If two points P and Q are not neighboring, we must choose some path C on the surface connecting P and Q, and break it up into small straight line segments PP', $P'P''$, ..., Q. If we move from P to P' along straight-line segment PP', we must rotate $T(P)$ at P through some small angle $\Delta\theta$ about the normal $\boldsymbol{n}$ at P in order to get the tangent plane $T(P')$ at P'. For the next segment $P'P''$, we have to rotate the tangent plane $T(P')$ through an angle $\Delta\theta'$ about the normal $\boldsymbol{n}'$ at P' in order to get the tangent plane $T(P'')$ at P''. We keep doing this until we reach the endpoint Q. Now we increase the number of intermediate points indefinitely, and take the limit of this process so that the broken straight line segments approach the curve. This defines the vector in $T(Q)$ that is parallel to one in $T(P)$ *with respect to the path* C.

Note that we must add the last qualification because, unless S is a developable surface, the resulting parallelism in general will be path dependent. We can see this by looking at a small parallelogram with sides PP', $P'Q$ and PP'', $P''Q$. Since $T(P')$ and $T(P'')$ are not in general parallel to each other, the correspondence between vectors in the tangent planes $T(P)$ and $T(Q)$ that is set up by going via $T(P')$ is not in general the same as the one we get by going via $T(P'')$.

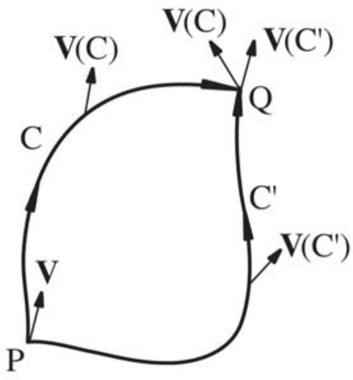

Figure 13: In general the vector $V'(C)$ at Q that is parallel to V at P depends on the path taken between P and Q.

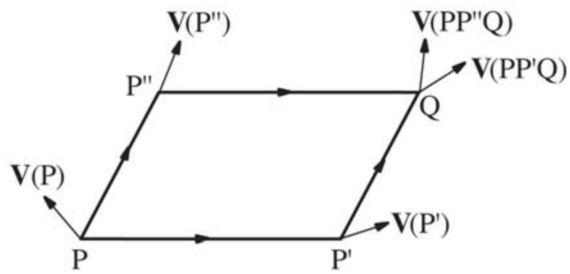

Figure 14: We can see this by looking at the parallel transport of a vector $\mathbf{v}$ along the sides PP', $P'Q$ and PP'', $P''Q$ of a small parallelogram.

d. the Riemann tensor. By carrying out the analysis of this parallelogram quantitatively, we can define the Riemann tensor of the surface.[47] Take a vector $\mathbf{v}$ in $T(P)$, and let the corresponding (i.e. intrinsically parallel) vector in $T(P')$ be $\mathbf{v} + \delta\mathbf{v}$. Then $\mathbf{v} + \delta\mathbf{v}$ results from $\mathbf{v}$ by a rotation operation that acts on $\mathbf{v}$; we shall symbolize it by the operator $\mathbf{I} + d\mathbf{O}$ (see eq. (14)), so that:

47 In the case of a two-dimensional surface, it reduces to a scalar R; i.e., all non-vanishing components of the Riemann tensor reduce to $\pm R$. But we prefer to keep the tensorial designation in view of the impending generalization to higher dimensions.

$$\delta \boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Here, $d\boldsymbol{\mathcal{O}}$ represents a first order infinitesimal rotation operator that depends linearly on $d\boldsymbol{r}$. Similarly, if $d\boldsymbol{r}'$ represents the displacement PP'', then the change $\delta\boldsymbol{v}'$ in $\boldsymbol{v}$ when we go from $T(P)$ to $T(P'')$ is given by:

$$\delta \boldsymbol{v}' = d\boldsymbol{\mathcal{O}}'\boldsymbol{v}.$$

Then the change in $\boldsymbol{v}$ at $T(Q)$ when we go via $d\boldsymbol{r}$ first, then $d\boldsymbol{r}'$ (i.e., via $PP'Q$) is given by:

$$\delta \boldsymbol{v}_1 = (\boldsymbol{I} + d\boldsymbol{\mathcal{O}}')(\boldsymbol{I} + d\boldsymbol{\mathcal{O}})\boldsymbol{v} - \boldsymbol{v} = (d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}'d\boldsymbol{\mathcal{O}}v)\boldsymbol{v};$$

while, if we proceed in the reverse order (i.e., via $PP''Q$), the change is given by:

$$\delta \boldsymbol{v}_2 = (d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}}d\boldsymbol{\mathcal{O}}')\boldsymbol{v}.$$

Since $\delta\boldsymbol{v}_1$ and $\delta\boldsymbol{v}_2$ are vectors at the same point, their difference is a (second order infinitesimal) vector $\delta^2\boldsymbol{v}$. It indicates by how much the two vectors in $T(Q)$ that are parallel to $\boldsymbol{v}$ in $T(P)$, depending on which of the two paths is taken, differ from each other:

$$\delta^2 \boldsymbol{v} = (d\boldsymbol{\mathcal{O}}d\boldsymbol{\mathcal{O}}' - d\boldsymbol{\mathcal{O}}'d\boldsymbol{\mathcal{O}})\boldsymbol{v}.$$

Note the operator in parentheses is the same for all vectors in $T(P)$ since they are all rotated by the same amount. And since $d\boldsymbol{\mathcal{O}}$ and $d\boldsymbol{\mathcal{O}}'$ are linear in $d\boldsymbol{r}, d\boldsymbol{r}'$, respectively, this operator is proportional to $(d\boldsymbol{r}d\boldsymbol{r}' - d\boldsymbol{r}'d\boldsymbol{r})$. Such an antisymmetric tensorial product of two vectors is abbreviated as $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and called a simple bivector; it represents the (signed) area of the infinitesimal parallelogram with sides $d\boldsymbol{r}, d\boldsymbol{r}'$. This second order infinitesimal term is also same for all vectors taken from P to Q along the sides of the parallelogram.[48] So there must be a finite tensorial operator $\boldsymbol{R}$, such that, when it operates on an area bivector $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and a vector $\boldsymbol{v}$, it produces the change in $\boldsymbol{v}$ when it is parallel transported around the area $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. Note that, to the second differential order we are considering, it makes no difference whether we parallel transport a vector from P to Q in two different ways, and compare the results in $T(Q)$, or take it around the parallelogram and compare the result with the original vector in $T(P)$. Further, the result is independent of the shape of the infinitesimal plane figure we carry it around so long as this has the same area as, and lies in the plane defined by, $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. The tensorial operator $\boldsymbol{R}$, which operates on a bivector and a vector to produce another vector, is called the *Riemann tensor*; when it operates on an infinitesimal area element, it measures how much Riemannian parallelism

48 One should actually distinguish between $d\boldsymbol{r}$ at P and $d\boldsymbol{r}$ at P'', which is the result of parallel transporting $d\boldsymbol{r}$ at P along $d\boldsymbol{r}'$. But to the order we are considering, the difference may be neglected. The more serious problem of whether the parallelogram resulting from these displacements actually "closes" will be discussed later.

on that surface element differs from flat, path-independent, parallelism, for which the Riemann tensor would vanish.

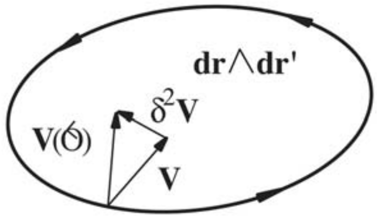

Figure 15: The operator $\boldsymbol{R}$, *operating on the area* $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ *and the vector* $\boldsymbol{v}$, *produces the change* $\delta^2 \boldsymbol{v}$ *in* $\boldsymbol{v}$ *when it is parallel transported around that area.*

e. non-flat affine spaces. Our discussions of parallelism on a surface and of the Riemann tensor made essential use of the metric of the enveloping Euclidean space. First of all, this metric induced a notion of distance on the surface; but this is intrinsic to the surface, and can be defined without using the fact that the surface is embedded in a Euclidean space. More serious is the fact that we used the normals to the surface at each point in order to develop the relation between tangent spaces at neighboring points in terms of an orthogonal transformation (rotation through some angle). The notions of orthogonality and angle are intrinsically metrical.

Suppose we abstract from these metric concepts and consider an affine space, as discussed above. Using only affine concepts, can we still define concepts of parallelism and straight line on a surface in an affine space? The answer is yes, but we must introduce a substitute for the unit normal field given naturally in a Euclidean space.

First of all, the concept of surface is independent of a metric, as are those of the tangent space at each point of a surface, and (hence) of the tangent bundle. But now we have no natural way of relating the tangent spaces at different point of the surface by means of a general linear transformation. At each point of the surface, a basis in its tangent space must be supplemented by a vector that does not lie in the tangent space; i.e., a vector that takes the place of the normal vector to the surface in a Euclidean space. Together with the chosen basis in the tangent space, this vector constitutes a basis for the enveloping affine space. This vector field is said to *rig* the surface, and the process is called *rigging*. Once the surface is rigged, one can carry out in an affine space a procedure to relate neighboring tangent spaces that is entirely analogous to the procedure used in the Euclidean case. The only difference is that, instead of the infinitesimal orthogonal transformation $d\mathcal{O}$ that carries the orthonormal basis at P into the orthonormal basis at P', one considers the infinitesimal general linear transformation $d\mathcal{L}$ that takes a basis for the enveloping affine space at P into the corresponding basis at P'. Due to the linearity of vector spaces, carrying out the transformation $d\mathcal{L}$ on any vector in $T(P)$ yields the corresponding parallel vector in $T(P')$. Such a connection between tangent spaces, which generalizes to surfaces in an affine space the Levi-Civita connection for surfaces in a Euclidean space, is called *a general linear connection*. Once the connection is defined, everything proceeds in a way that is entirely analogous to that for Euclidean spaces (see the previous subsection), up to and including the definition of the Riemann tensor operator.

Instead of eq. (15), giving the effect of an infinitesimal orthogonal transformation (rotation) matrix on an orthonormal basis, we can now specify the effect on the vectors in a tangent plane of an infinitesimal general linear transformation, by specifying the infinitesimal general linear transformation matrix $(dL)^A_B$ that gives the effect of this transformation on an arbitrary basis:

$$\delta \boldsymbol{e}_B(P') = (dL)^A_B \boldsymbol{e}_A. \tag{16}$$

f. covariant differentiation, geodesics. Once we have the concept of parallelism along a path, we can define a derivative operation for a vector field on a surface. The essence of the usual derivative operation for a vector field in Euclidean space consists in comparing the value of the vector field $\boldsymbol{v}$ at some point with its values at some neighboring points. But we can only compare vectors in the same tangent space: what we actually do to compare vectors at two points P and Q is to compare $v(Q)$ with the vector at Q that is parallel to $\boldsymbol{v}(P)$. We shall proceed in the same way on a surface and compare values at two neighboring points P and $P + d\boldsymbol{r}$:

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta\boldsymbol{v}] = d\boldsymbol{r} \cdot \partial\boldsymbol{v} - \delta\boldsymbol{v},$$

since

$$\boldsymbol{v}(P + d\boldsymbol{r}) = \boldsymbol{v}(P) + d\boldsymbol{r} \cdot \partial\boldsymbol{v},$$

where ∂ represents the ordinary-derivative gradient operation; operating on a scalar field $\phi(\boldsymbol{r})$, it gives the gradient vector field $\partial\phi$; but operating on a vector (or tensor) field it does not produce another vector (or tensor) field. It must be supplemented by the second term $\delta\boldsymbol{v}$ for a vector (and similar terms for higher-order tensors). Since $\delta\boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}$, we can write the invariant combination as

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta\boldsymbol{v}] = d\boldsymbol{r} \cdot \partial\boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Since $d\boldsymbol{\mathcal{O}}\boldsymbol{v}$ is also linear in $d\boldsymbol{r}$, we can abbreviate the right hand side as:

$$d\boldsymbol{r} \cdot \partial\boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v} = d\boldsymbol{r} \cdot \nabla\boldsymbol{v}.$$

The expression $d\boldsymbol{r} \cdot \nabla$ represents an invariant directional derivative in the $d\boldsymbol{r}$ direction. Since the result is linear in $d\boldsymbol{r}$, there must be a tensorial operator $\boldsymbol{\nabla}$ called the *covariant derivative* operator, that operates on a vector to produce a mixed tensor $\boldsymbol{\nabla v}$ with one covariant and one contravariant (i.e.,vectorial) place.

On a surface, we may generalize the concept of a straight line in an affine space to that of a geodesic by requiring that the parallel transport of its tangent vector along a geodesic remain the tangent vector. If $\boldsymbol{t}$ represents the tangent vector to the curve $C(\lambda)$, $\boldsymbol{t} = dC/d\lambda$, this means that a geodesic must satisfy the equation:

$$\boldsymbol{t} \cdot \nabla \boldsymbol{t} = 0.$$

g. generalizations, intrinsic characterizations. Nothing in the discussion above depends essentially on the number of dimensions being three, and it can be immediately generalized to n-dimensional metric and affine spaces, defined by the translation groups $T(n)$ and $O(n, R)$ and $T(n)$ and $GL(n, R)$ respectively; and to their m-dimensional sub-spaces. If m is less than $n-1$, then there are n-m normals, and n-m rigging vectors must be defined; but otherwise the discussion proceeds quite analogously. Since any m-dimensional Riemannian or affinely-connected space can be embedded in an n-dimensional Euclidean or affine space of sufficiently high dimension (locally, if not globally), such embedding arguments can handle the generic case.

Of course, once the basic geometrical concepts have been grasped, an intrinsic method of characterizing curved spaces, independently of any embedding in flat spaces of higher dimension, is preferable. It is clear from the previous discussion how to proceed. One must specify a connection between vectors in $T(P)$ and $T(P')$ that defines when a vector in one is parallel to a vector in the other. In contrast to the order in the previous embedding considerations, I shall first give the definition for a general affine linear connection, and then indicate how to specialize it to a Riemannian or Levi-Civita connection.

As indicated earlier (see discussion around eqs. (15) and (16) above), in order to connect arbitrary vectors in the two tangent spaces, it suffices to indicate how sets of basis vectors in the two tangent spaces are connected. Let $\boldsymbol{e}_i(P)$ be a set of basis vectors in $T(P)(i = 1, 2, \ldots, n)$. The changes in these basis vectors when we move to $T(P + d\boldsymbol{r})$ will be given by (generalizing eq. (6) above):

$$\delta \boldsymbol{e}_i(P') = (dL)_i^j \, \boldsymbol{e}_j, \qquad (i, j = 1, 2, \ldots, n).$$

Our connection is linear in $d\boldsymbol{r}$, so it suffices to know the change in $\boldsymbol{e}_i$ for a small change in each of the basis directions, $d\boldsymbol{r} = \epsilon e_k$, where ϵ is an infinitesimal of first order.

$$\delta \boldsymbol{e}_i(P') = \epsilon L_i^j(P, \boldsymbol{e}_k) \boldsymbol{e}_j.$$

On the other hand $\delta \boldsymbol{e}_i$ itself must be a linear combination of the basis vectors, so we may decompose it into the infinitesimal changes in each of these directions:

$$(\delta \boldsymbol{e}_i)_j = \epsilon \Gamma_{ij}^k \boldsymbol{e}_k.$$

Thus, specification of the set of quantities $\Gamma_{ij}^k(P)$ at all points of the manifold fixes the affine connection intrinsically.[49] We call the Γ_{ij}^k the *components of the connection* with respect to the basis $\boldsymbol{e}_i$.[50]

If we now want to construct the parallelogram as described above in the definition of the Riemann tensor, we must make sure that it "closes," that is, that we reach the same point if we parallel transport $d\boldsymbol{r}$ along $d\boldsymbol{r}'$ as we do if we parallel transport $d\boldsymbol{r}'$

along $d\boldsymbol{r}$. It is relatively simple to show that this will be the case if Γ^k_{ij} is symmetric in its two lower indices; we shall consider only such symmetric affine connections.[51]

The Riemann tensor operator $\boldsymbol{R}$ can now be defined in terms of its effect on the basis vectors. If we transport $\boldsymbol{e}_k$ around an area defined by $\boldsymbol{e}_i \wedge \boldsymbol{e}_j$, then its change in the $\boldsymbol{e}_l$ direction is given by R^k_{ijl}. These are the components of the Riemann tensor with respect to the basis $\boldsymbol{e}_i$, which can easily be related to the derivatives of the Γ^j_{ki}, but we omit the details. For future reference, we note that R^k_{ijl} is antisymmetric in its last pair of indices, and that if we contract its upper index with either of the last two indices, say the second, we get (plus or minus) the Ricci tensor R_{il}.

The covariant derivative operator will have components:

$$\nabla_i = \boldsymbol{e}_i \cdot \nabla;$$

the components of the covariant derivative of a vector $\boldsymbol{\nabla v}$, for example, are:

$$\nabla_i v^j = \partial_i v^j + \Gamma^j_{ki} v^k.$$

The components of the geodesic equation in an adapted coordinate system are:

$$d^2x^j/d\lambda^2 + \Gamma^j_{ki}(dx^k/d\lambda)(dx^i/d\lambda) = 0.$$

The components of the Riemann tensor with respect to a basis can be similarly calculated.

Turning to Riemannian spaces, it is natural to demand that parallel transport along any path preserve the length of all vectors. If we impose this condition on a symmetric affine connection, we are led uniquely to the Levi-Civita connection discussed above; but again we omit the details.

For future reference, we also note that, just as in the case of a surface in a linear (flat) affine space discussed above, a connection is induced on a hypersurface in a non-flat affinely connected space if that hypersurface is rigged with an arbitrary vector field.[52]

---

49 Note that these quantities transform as scalars under a coordinate transformation, but as tensors under a change of basis. If we use the natural basis associated with a coordinate system (see the following note) and carry out a simultaneous coordinate transformation and change of natural basis, they transform under a more complicated, non-tensorial transformation law (see Section 6, eq. (7)).

50 Note that a basis need not be holonomic, i.e., coordinate forming. It will be if and only if the Lie bracket of any pair of basis vectors vanishes. We shall only need holonomic bases, for which an associated coordinate system x^i exists, such that in this coordinate system e^j_i, the coordinate components of $\boldsymbol{e}_i$, are equal to δ^j_i, the Kronecker delta. Conversely, a basis is associated with any coordinate system by the same relations.

51 If the parallelogram does not close, the antisymmetric part of Γ^j_{ki} defines the so-called *torsion tensor*.

52 It is customary, when discussing spaces of more than three dimensions, to refer to subspaces of one less dimension than that of the space as *hypersurfaces*. Thus, when the discussion is generalized to more than three dimensions, "surfaces" become "hypersurfaces."

h. frame bundles and connections. We introduced the concept of affine connection in the currently-habitual way, in terms of its local action on vector or frame fields in some manifold. But a connection is more naturally introduced globally in terms of the frame bundle over that base manifold (see Section b). A curve C in the base manifold together with a frame field defined along the curve corresponds to a curve $\mathbf{C}$ in the frame bundle; and conversely $\mathbf{C}$ projects down to C in the base manifold, together with a frame field along the curve. Now a connection provides a rule for defining such curves in the frame bundle: given a curve C in the base manifold together with an initial frame at some point on the curve, parallel transport of the initial frame along the curve thus defines a unique curve $\mathbf{C}$ in the frame bundle. The only thing we have to worry about is what happens if we change the initial frame by the action of some element $\mathbf{L}^i_j$ of the general linear group (see eq. (11)).The curve in the frame bundle is then transformed into another curve that differs from the first only by the same action of $\mathbf{L}^i_j$ on the frame at each point of the curve in the base manifold.

We can use this idea to define a connection globally as a collection of curves in the frame bundle, each passing only once through any fibre of the bundle, that satisfy the following condition: if two such curves $\mathbf{C}$ and $\mathbf{C}'$ project into the same curve C in the base manifold, and hence have all of their fibres in common, then on each fibre the frames on the two curves are related by the global application of the same $\mathbf{L}^i_j$.

If we want to restrict the structure group of the frame fibres to some subgroup of $GL(n, r)$, then we must assure that the connection introduced is compatible with the structure of this subgroup. For example, if we required compatibility with any of the orthogonal or pseudo-orthogonal subgroups, the Levi-Civita connection would result.[53]

REFERENCES

Bhaskar, Roy. 1993. *Dialectic: The Pulse of Freedom*. London/New York: Verso.

Bonola, Roberto. 1955. *Non-Euclidean Geometry: A Critical and Historical Study of its Development*. Dover reprint: New York: Dover Publications. (Open Court 1912)

Cartan, Élie. 1923. "Sur les variétés à connection affine et la théorie de la relativité généralisée." *Ecole Normale Supérieure* (Paris). *Annales* 40: 325–412. English translation in (Cartan 1986).

———. 1986. *On Manifolds with an Affine Connection and the Theory of Relativity.* Translated by Anne Magnon and Abhay Ashtekar. Naples: Bibliopolis.

Collier, Andrew. 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy.* London/New York: Verso.

Coolidge, Julian Lowell. 1940. *A History of Geometrical Methods*. Oxford: Clarendon Press.

Crowe, Michael J. 1994. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. New York: Dover Publications (reprint ed.).

Crampin, Michael, and Felix A.E. Pirani. 1986. *Applicable Differential Geometry.* Cambridge/London/ New York: Cambridge University Press.

Darling, R. W. 1994. *Differential Forms and Connections*. Cambridge/New York/Melbourne: Cambridge University Press.

Ehlers, Jürgen. 1981. "Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie." In J. Nitsch et al. (eds.), *Grundlagenprobleme der modernen Physik*. Mannheim: Bibliographisches Institut, 65–84.

53 See (Crampin and Pirani 1986, chap. 15) for details.

Einstein, Albert. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen folgerungen." *Jahrbuch der Radioaktivität und Elektronik*. 4: 411–462.

———. 1914. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse. Sitzungsberichte*: 1030–1085.

———. 1915. "Zur allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse. Sitzungsberichte*: 778–786.

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. I. Physikalischer Teil von Albert Einstein. II. Mathematischer Teil von Marcel Grossmann*. Leipzig: Teubner.

Eisenstaedt, Jean and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies*, vol 3). Boston/Basel/Berlin: Birkhäuser.

Friedrichs, K. 1927. "Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz." *Mathematische Annalen* 98: 566–575.

Gauss; Carl Friedrich. 1902. *General Investigations of Curved Surfaces*. Princeton: Princeton University Press. English translation by A. Hiltebeitel and J. Morehead of *Disquisitiones generales circa superficies curvas* (Göttingen: Dietrich, 1828).

Grassmann, Hermann. 1844. *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*. Leipzig: Otto Wigand.

———. 1862. *Die Ausdehnungslehre. Vollständig und in strenger Form bearbeitet*. Berlin: Enslin.

———. 1877. "Ueber die Beziehung der nicht-Euklidischen Geometrie zur Ausdehnungslehre."Appendix I to (Grossmann 1878). English translation in (Grossmann 1995, 279–280).

———. 1878. *Die Ausdehnungslehre von 1844 oder Die lineale Ausdehnungslehre*. (2nd. Edition.) Leipzig: Otto Wigand.

———. 1995. *A New Branch of Mathematics: The Ausdehnungslehre of 1844 and Other Works*. Translated by Lloyd C. Kannenberg. Chicago and LaSalle: Open Court.

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik. In neuem Zusammenhänge dargestellt*, Philipp Lenard (ed.). *Gesammelte Werke*, Vol. 3. Leipzig: Barth.

Hessenberg, Gerhard. 1917. "Vektorielle Begründung der Differentialgeometrie." *Mathematische Annalen* 78: 187–217.

Howard, Don and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston/Basel/Berlin: Birkhäuser.

Janssen, Michel. 1992. *H. A. Lorentz's Attempt to Give a Coordinate-Free Formulation of the General Theory of Relativity*." In (Eisenstaedt and Kox 1992, 344–363).

Jost, Jürgen. 1991. *Riemannian Geometry and Geometric Analysis*, 2nd. ed. Berlin/Heidelberg/New York: Springer.

Klein, Felix. 1872. *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Erlangen: A. Dühechert. Revised version in *Mathematische Annalen* 43: 63–100 (1893).

Kolmogorov, A. N. and A. P. Yushkevich. 1996. *Mathematics of the 19th Century*. Basel/Boston/Berlin: Birkhäuser.

Laptev, B. L. and B. A. Rozenfel'd. 1996. "Chapter 1. Geometry." In (Kolmogorov and Yushkevich 1996, 1–118).

Lawvere, F. William. 1996. "Grassmann's Dialectics and Category Theory." In (Schubring 1996, 255–264).

Levi-Civita, Tullio. 1916. "Nozione de Parallelismo in una Varieta Qualunque e Conseguente Specificazione Geometrica della Curvatura Riemanniana." *Rendiconti del Circolo Matematico di Palermo* 42: 17–205.

Lützen, Jesper. 1995a. "Interactions Between Mechanics and Differential Geometry in the 19th Century." *Archive for History of Exact Sciences* 49: 1–72.

———. 1995b. "Renouncing Forces: Geometrizing Mechanics. Hertz's Principles of Mechanics." *Københavns Universitet Matematisk Institut, Preprint Series* 1995, No. 22.

Norton, John. 1984. "How Einstein Found his Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316.

Reich, Karen. 1992. Levi -Civitasche Parallelverschiebung, affiner Zusammenhang, Übertragungsprinzip: 1916/17–1922/23." *Archive for History of Exact Sciences* 44: 78–105.

———. 1994. *Die Entwicklung des Tensorkalküls. Vom absoluten Differentialkalkuel zur Relativitätstheorie*. Basel: Birkhäuser.

Riemann, Bernhard. 1868. "Über die Hypothesen, welche die Geometrie zu Grunde liegen." In Richard Dedekind and H. Weber (eds.), *Gesammelte mathematische Werke*, 2nd ed. 1892, (reprinted in New York: Dover Publications, 1953), 272–287.

Scholz, Erhard. 1995. "Hermann Weyl's Purely Infinitesimal Geometry." In *Proceedings of the International Congress of Mathematicians, Zurich, Switzerland 1994*. Basel: Birkhäuser.

Schouten, J. A. 1918. “Die direkte Analysis zur neueren Relativitätstheorie.” *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, XII, no. 6.

Schubring, Gert (ed.). 1996. *Hermann Guenther Grassmann (1809–1877): Visionary Mathematician, Scientist and Neohumanist Scholar*. Dordrecht/Boston/London: Kluwer Academic.

Stachel, John. 1986. “What a Physicist Can Learn from the Discovery of General Relativity.” In Remo Ruffini (ed.), *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity.* Amsterdam: Elsevier, 1857–1862.

———. 1987. “How Einstein Discovered General Relativity: A Historical Tale With Some Contemporary Morals.” In Malcolm A. H. MacCallum (ed.), *General Relativity and Gravitation: Proceedings of the 11th International Conference on General Relativity and Gravitation*. Cambridge: Cambridge University Press, 200–208.

———. 1989. “Einstein’s Search for General Covariance, 1912–1915.” In (Howard and Stachel 1989, 63–100).

———. 1993. “The Meaning of General Covariance: The Hole Story.” In John Earman et al. (eds.), *Philosophical Problems of the Internal and External World, Essays on the Philosophy of Adolf Grünbaum*. (Konstanz: Universitätsverlag, Pittsburgh University Press, 129–160.

———. 1994a. “Scientific Discoveries as Historical Artifacts.” In Kostas Gavroglu (ed.), *Current Trends in the Historiography of Science*. Dordrecht: Reidel, 139–148.

———. 1994b. “Changes in the Concepts of Space and Time Brought About by Relativity.” In Carol C. Gould and Robert S. Cohen (eds.), *Artefacts, Representations and Social Practice*. Dordrecht: Kluwer, 141–162.

———. 1995. “History of Relativity.” In Laurie M. Brown, Abraham Pais and Brian Pippard (eds.), *Twentieth Century Physics*, Vol. I. Bristol and Phila.: Institute of Physics Pub., New York: American Institute of Physics Press.

——— (ed.). 1998. *Einstein’s Miraculous Year: Five Papers That Changed the Face of Physics*. Princeton: Princeton University Press.

———. 2003a. “Critical Realism: Bhaskar and Wartofsky.” In Carol C. Gould (ed.), *Constructivism and Practice: Towards a Historical Epistemology.* Lanham, Md.: Rowman and Littlefield, 137–150.

———. 2003b. “Einstein’s Intuition and the Post-Newtonian Approximation.” Talk at the Mexico City Conference in Honor of Jerzy Plebanski, May 26, 2003. To appear in the Proceedings of the Conference.

———. 2005. “Fibered Manifolds, Geometric Objects, Structured Sets, G-Spaces and All That: The Hole Story from Space-Time to Elementary Particles.” To appear in (Stachel forthcoming).

———. Forthcoming. *Going Critical: Selected Essays*. Dordrecht: Kluwer.

Struik, Dirk. 1933. “Outline of the History of Differential Geometry.” *Isis* 19: 92–120, *Isis* 20: 161–191.

Weyl, Hermann. 1918a. “Reine Infinitesimalgeometrie.” *Mathematische Annalen* 2: 384–411.

———. 1918b. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. Berlin: Springer.

———. 1919. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 3rd ed.

———. 1921. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 4th ed.

———. 1923. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 5th ed.

Zuckerman, Harriet and Joshua Lederberg. 1986. Forty years of genetic recombination in bacteria: postmature scientific discovery?” *Nature* 324: 629–631.