[19]A particularly clear exposition of Jarrett's argument has been given by L. Ballentine and J. Jarrett, "Bell's theorem: Does quantum mechanics contradict relativity?" Am. J. Phys. 55, 696–701 (1987).

[20]Your students need not even have learned about bra vectors or inner products. They need only solve Eq. (17) to express the single-particle 1 states in terms of the 2 states, make the appropriate substitutions into Eq. (16), and read off the squared coefficients.

[21]Abner Shimony calls it "outcome independence." I prefer Jarrett's terminology.

[22]Shimony calls it "parameter independence." I again prefer Jarrett's terminology.

[23]The last paragraph of Appendix A illustrates the fact that when averaged over λ Eq. (B4) does indeed hold in a Hardy state.

[24]A more graceful but more subtle (though entirely correct) route from Eq. (B5) to Eq. (11) consists of simply noting that $\langle p_\lambda^l(2X')p_\lambda^l(1X)p_\lambda^r(1Y)p_\lambda^r(2Y')\rangle$ can be interpreted as a distribution for an ensemble of pairs of particles in which each member of the pair has a specified outcome (R or G) for each of the switch settings (1 or 2) it might encounter, and in which the marginal distributions that describe each of the four sets of experiments one might actually perform (11, 12, 21, or 22) agree with the experimental distributions. If the experimental distributions can indeed be simulated by such an ensemble, then my derivation of Eq. (11) in Sec. IV in the manner of Stapp is indisputably valid, whether or not that ensemble makes any physical sense.

General relativity before special relativity: An unconventional overview of relativity theory

Wolfgang Rindler
Physics Department, The University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 7 March 1994; accepted 5 April 1994)

It is suggested how Bernhard Riemann might have discovered General Relativity soon after 1854 and how today's undergraduate students can be given a glimpse of this before, or independently of, their study of Special Relativity. At the same time, the whole field of relativity theory is briefly surveyed from the space–time point of view.

I. INTRODUCTION

Historically, Einstein's General Relativity of 1915—the theory of curved spacetime—arose as a generalization of his Special Relativity of 1905—the theory of flat spacetime—much as the geometry of curved surfaces arises as a generalization of the Euclidean geometry of the plane. This historical sequence from the special to the general theory is followed in every presentation of the subject known to me. And for good reason: in this way the required level of mathematical sophistication rises only gradually, whereas the inverse sequence would seem to require some heavy mathematics up front. However, it is amusing and instructive to fantasize how, in the best of all possible worlds, General Relativity might have been developed *ab initio* long before 1905, for example by Bernhard Riemann soon after 1854, and how it could then have led to Special Relativity. At the same time, a mathematically diluted version of such a development can prove to be of interest to bright undergraduate students. It gives them a quick and direct taste of spacetime and of General Relativity, two topics which are often promised them "at the end of Special Relativity," but which only too often are never quite reached. This sequence also well illuminates the inner logic and self-sufficiency of General Relativity.

The following is a sample of such a development, which, with suitable omissions, can be presented to students in an hour's lecture.

II. HOW THEORIES ORIGINATE

New theories are as a rule not developed for sport. Rather, they arise in response to difficulties, paradoxes, or puzzles in the older theories. Thus Special Relativity grew out of difficulties in reconciling Maxwell's theory with Newtonian kinematics, and, in spite of Einstein's well-known disclaimer, it could hardly have come into being without the acute paradox of the Michelson–Morley experiment of 1887. This experiment showed that, no matter how fast you chase a light signal, you can never reduce its speed relative to you. General Relativity, on the other hand, has its roots in the much older mechanics of Newton. But Newton's theory, too, is by no means free of puzzles. Above all, it has long been criticized for its reliance, if not necessarily on absolute space, on the set of global inertial frames whose absoluteness ("they act but cannot be acted on") so offended the scientific sensitivities of Mach and Einstein. And then there is the mystery of the equality of gravitational and inertial mass, appearing simply as a postulate in Newton's theory. Why should a quantity measuring a body's inertia or resistance to acceleration act at the same time as its "gravitational charge?" It would seem that these two puzzles alone (and there were others) could drive a man to search for a new theory, i.e., a new mathematical model, especially when a new and suitable mathematical avenue had just opened up. The man might have been Riemann, and the avenue his newly discovered differential geometry of (irregularly) curved spaces of higher dimensions.

III. GAUSS' GEOMETRY OF SURFACES

The year 1854 was a memorable one in the annals of the famous old German university town of Göttingen. The recently developed railroad had finally reached the town. And also, though unbeknown to most of its good burghers, the

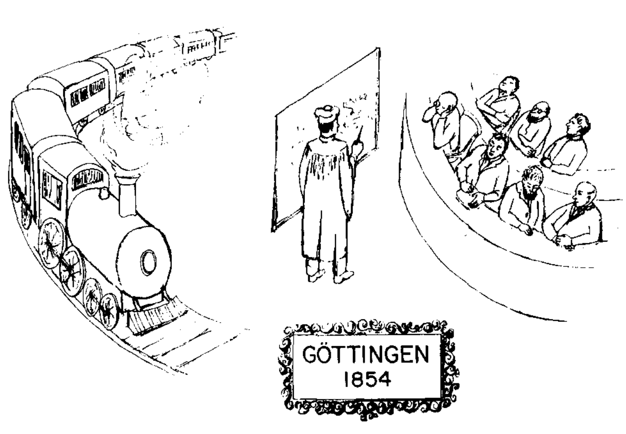

Fig. 1. In 1854 the railroad reached Göttingen, but Bernhard Riemann unveiled there an even more important system of rails, the geodesics of curved spaces.

28-year-old Bernhard Riemann unveiled to his "Habilitation" examiners the outlines of his groundbreaking new geometry.[1] This was in time to furnish the universe with a different and more permanent network of rails: the "geodesics" along which free particles are constrained to move in curved space–time. (See Fig. 1.)

Riemann's teacher, the great Gauss, had already put the differential geometry of *two*-dimensional surfaces (i.e., of the surfaces we know from everyday life) on a firm basis. As Gauss stressed, the inner or *intrinsic* differential geometry of a surface is completely determined by its *metric*, i.e., by the formula giving the distance ds between any two neighboring points on it. For example, for a sphere of radius a (see Fig. 2) we can write

$$ds^2=a^2(d\theta^2+\sin^2\theta d\phi^2), \tag{1}$$

where θ and ϕ are the usual angles of "colatitude" and longitude, respectively. Even if I had never seen a sphere in my life, I could from this concise "blueprint" construct one. I might start by making a flat map of the surface, labeling the vertical and horizontal lines of some arbitrary rectangular grid θ=0, 0.1, 0.2,... and ϕ=0, 0.1, 0.2,..., say. Then, using the formula (1), I could write in the actual lengths corresponding to the sides of the elementary squares of my grid, as well as of one of the diagonals of each square. Then I could cut out from cardboard little triangles having these actual dimensions (two for each square of the grid), glue them together according to the map, and the result would be the sphere!

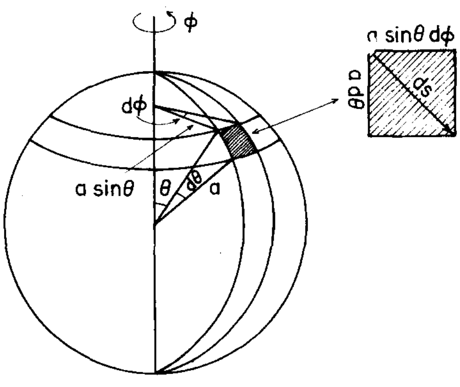

Fig. 2. A coordinate displacement $d\theta$, $d\phi$ on a sphere produces a distance displacement ds given by $ds^2=a^2(d\theta^2+\sin^2\theta d\phi^2)$.

Gauss emphasized the irrelevance of the *particular* coordinates chosen. One can quite arbitrarily draw on any surface two families of mutually intersecting curves, label them arbitrarily u=0, 0.1, 0.2,... and v=0, 0.1, 0.2,..., say, and then express ds^2 in terms of these "Gaussian" coordinates u and v:

$$ds^2=Edu^2+2Fdudv+Gdv^2, \tag{2}$$

where E,F,G will in general be functions of u and v. The reason why the metric ds^2 will always be a *quadratic* in the coordinate differentials du,dv is that the surface is a subspace of our everyday Euclidean three-space E_3. The metric of this latter space, when referred to Cartesian coordinates x,y,z, is

$$ds^2=dx^2+dy^2+dz^2. \tag{3}$$

And since every point on the surface is also a point of E_3, we must have three relations of the form $x=f(u,v)$, $y=g(u,v)$, $z=h(u,v)$, which, when substituted into Eq. (3), lead to a quadratic form like Eq. (2).

A surface corresponding to a given metric might well be deformable without stretching or tearing, e.g., a plane into a cone or a cylinder. *Intrinsic* geometric properties are those depending *only* on the metric and they are thus preserved under any such bending. The most fundamental intrinsic structure after the metric on any surface is the totality of its *geodesics*, these being the analogues of straight lines in the plane. A geodesic can be defined as the "straightest" path on a curved surface. For example, on a sphere the geodesics are all the great circles. If I just follow my nose, I will walk along a great circle. If I draw a straight line down the middle of a length of Scotch tape, and then carefully glue that tape inch by inch without wrinkles onto the surface—any surface—the center line I have drawn will be a geodesic on that surface. Alternatively, but equivalently, a geodesic can be defined by the property that every sufficiently short portion AB of it represents the shortest path between A and B. In a notation useful for our later purposes, this property can be written symbolically in the form

$$\delta\int_A^B ds=0, \tag{4}$$

which states that the "first variation" of the distance $\int_A^B ds$ vanishes.

The intuitive expectation is that there is exactly *one* geodesic through any given point on the surface with given initial direction, and this can indeed be established as a theorem. It will play a *key role* in our development.

Geodesics serve among other things to define the ("Gaussian") curvature K of the surface at a given point P. Draw two neighboring geodesics through P and let their (small) separation η be a function of the distance s from P. Then K is defined by the equation

$$K=-\lim_{s\to 0}\left(\frac{d^2\eta}{ds^2}\Big/\eta\right) \tag{5}$$

and Gauss proved that *any* neighboring pair of geodesics through P yields the same K. On the sphere (1) one finds $K=1/a^2$ (exercise!). The sphere has constant curvature, but for a general surface K varies from point to point.

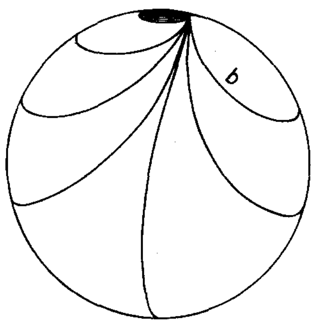

Fig. 3. A flatman F blows up a one-dimensional balloon b (a circle) on a sphere. Eventually the balloon envelops him.

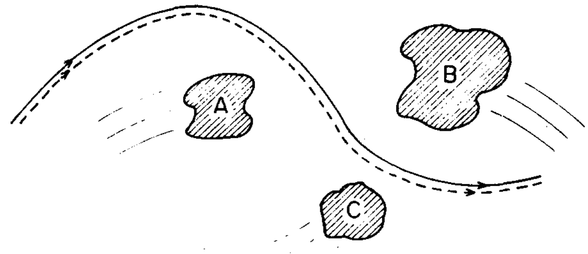

Fig. 4. A piano (full-line path) and a ping-pong ball (dotted-line path) fly side by side through the gravitational field caused by some gravitating bodies A,B,C.

IV. RIEMANNIAN GEOMETRY

Gauss' differential geometry for two-surfaces was generalized by Riemann to n dimensions in what may today seem an almost straightforward manner. But the very concept of irregularly curved spaces of higher dimensions was new and daring at the time. Just as the blueprint for a two-surface is encapsuled into a metric of the form (2), that for a "Riemannian" space of n dimensions, referred to Gaussian (i.e., arbitrary) coordinates $x_1,x_2,\ldots,x_n$, is encapsuled into a similar (i.e., quadratic) n-dimensional metric of the form

$$ds^2=\sum g_{ij}dx_i dx_j, \quad (g_{ij}=g_{ji}). \tag{6}$$

There are three metric coefficients g_{ij} for a two-space [comparison with Eq. (2) shows $g_{11}=E$, $g_{12}=F$, $g_{22}=G$ if we set $x_1=u$ and $x_2=v$], six for a three-space, 10 for a four-space, and so on. All the g_{ij} are functions of position, i.e., functions of the coordinates x_i.

Again the most fundamental intrinsic structure of a Riemannian space, after its metric, is the totality of its geodesics. These are defined, just as in the two-dimensional case, as straightest, or alternatively, as shortest lines between sufficiently close points on them. And, most importantly, there is again a unique geodesic through a given point in a given direction. Curvature is defined as in the two-dimensional case, though now it is no longer a single number at a given point P. Rather, we must choose a "planar direction" (i.e., an infinitesimal plane element) at P, and in this planar direction choose two geodesics issuing from P, to which we then apply formula (5). In general, the curvature we find depends on the planar direction we choose; but these various curvatures are interrelated, in fact they constitute a "tensor," the so-called *Riemann curvature tensor*. Only in *spaces of constant curvature* is that curvature independent of planar direction and position. A three-space of constant curvature $K=1/a^2$ is called a three-sphere, and many of its properties are analogous to those of a two-sphere. Moreover there is a good chance that the actual universe we inhabit is just such a three-sphere with presently expanding "radius" a. Looking in any direction we all could, in principle, see the backs of our heads. If I blow up an infinitely elastic balloon in such a universe, there comes a moment when the balloon (always a two-sphere) attains maximum surface area, at which moment its inside and its outside are equal halves of the universe! And if I continue to blow, I ultimately find myself inside a *shrinking* balloon. For a two-dimensional analogy, imagine a "flatman" blowing up circles on a two-sphere! (See Fig. 3.)

V. NEWTON'S GRAVITATIONAL THEORY

Let us next look at the state of health of Newtonian gravitational theory at the time of Riemann. During the 18th century this theory had seen a tremendous flowering, owing mainly to the work of mathematicians like Euler, Lagrange, and Laplace, and to advances in the accuracy and in the volume of astronomical observations. With astonishing precision the mathematicians were able to explain each new irregularity in the motions of earth, moon, sun, and the planets discovered by the astronomers. Newton's theory had become, in Penrose's classification of theories into superb, useful, and merely tentative, a *superb* theory.

And yet it was not perfect. As we noted earlier, it was open to attack on one of its most indispensable tenets, the apparently god-given absolute nature of the inertial frames. Then there was the puzzle of the gravitational charge: if I have a large ball carrying an electric charge Q, a test particle of charge $-q$ at distance r will be attracted to the ball with a (Coulomb-) force $F=Qq/r^2$. By Newton's second law, $F=ma$, it will therefore fall towards the ball with an acceleration $a=Qq/mr^2$, m being the *inertial* mass of the particle. The acceleration thus depends on the ratio q/m and can be large or small or even zero. This is not so in the analogous gravitational case! Writing M' and m' now for the "gravitational charge" of the ball and of the particle, respectively, we have $F=GM'm'/r^2$ (G being Newton's constant of gravitation) and $a=GM'm'/mr^2$. As was first demonstrated on the Leaning Tower of Pisa by Galileo, all particles fall alike! The reason: $m'=m$, i.e., gravitational charge equals inertial mass. This extraordinary experimental fact is simply taken as an axiom into the theory. One of its consequences is what has been called *Galileo's principle*. This states that in the gravitational field of an arbitrary mass distribution the path in space *and* time of a test particle is determined fully by its initial velocity. If I project into this field a piano and a ping-pong ball side by side with the same initial velocity, they will stay side by side forever. (See Fig. 4.)

The absoluteness of inertial frames and the unexplained identity of gravitational charge with inertial mass can perhaps be regarded as "mere" philosophical difficulties. Gradually, however, there came to light also a certain small *numerical* discrepancy that would not go away. Work on the fine details of planetary motions had already led the French astronomer Leverrier in 1846—purely by calculation—to predict the existence of a new planet he called Neptune, which was found almost at once just where he predicted. By 1859 the same Leverrier had shown that of the observed

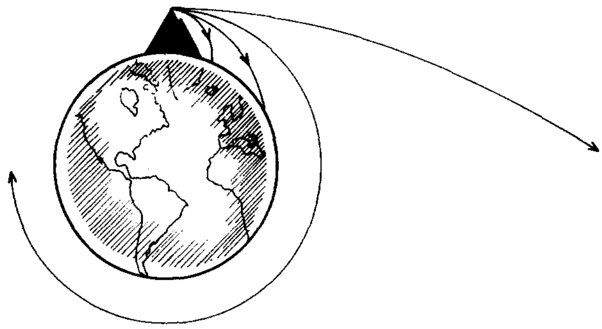

Fig. 5. The paths of some missiles shot off from a mountain top in the same horizontal direction but with different speeds.

precession of the orbit of Mercury only a part could be explained by Newtonian theory as due to other heavenly bodies. The missing part, some 40 arcsec per century (!!), was soon perceived as an acute puzzle, which was, in fact, not resolved until Einstein's General Relativity amazingly yielded the exact value (giving Einstein as a result several nights of sleeplessness from sheer excitement.)

As a fourth difficulty of Newtonian theory we can briefly mention its apparent nonapplicability to the dynamics of an infinite homogeneous universe of the kind contemplated then and, among other possibilities, even today. The problem is that the Newtonian forces in all directions are infinite (though balancing) and that no sensible potential can be found.

VI. A SCENARIO FOR DISCOVERING GENERAL RELATIVITY

So far we have sketched some parts of the scientific scene as it might have been present in Riemann's mind around 1860. The full extent of Maxwell's theory is still four years away and Einstein will not be born for another 19 years; Riemann has six more years to live. We now indulge in a game of "what if."

What if Riemann, contemplating the mystery of Galileo's principle (Fig. 4), had decided that it tells us something not so much about mass as about space? Space seems to have *rails* along which all "free" test particles (subject to no forces other than gravity) must travel alike. But there is just *one* set of natural rails in any space: its geodesics. Since Galileo's rails are patently curved, it would seem that space, too, would have to be curved. Could our three-space, then, be appropriately curved by the gravitating masses in it for its geodesics to provide the Galilean paths? The immediate answer is no, and a diagram (Fig. 5) essentially dating back to Newton's Pricipia well illustrates the reason. If we project a series of missiles from a mountain top, all horizontally and all in the same direction but with different speeds, some will drop nearby, the faster ones will go farther, some will orbit the earth, and some will escape to infinity. Evidently not all these paths can be geodesics in three-space, since the geodesic in a given direction is unique. Now Galileo's principle actually asserts that the path is fully determined by an initial *velocity* $(dx/dt,\ dy/dt,\ dz/dt)$. Knowing such a velocity in three-space, however, is equivalent to knowing a *direction* $dx{:}dy{:}dz{:}dt$ in space and time. Moreover, the piano and the ping-pong ball stay together not only in space but also in time. So could it be that the orbits are geodesics in "space and time?" One can easily contemplate a four-dimensional spacetime whose "points" are all the *events* (x,y,z,t) of this world. But what would be its metric? We cannot simply add something like dt^2 to something like $dx^2+dy^2+dz^2$, since a sum of square centimeters and square seconds makes no sense. At least we must multiply dt^2 by a factor having the dimensions of a velocity squared. Since we are dealing with gravity, such a factor readily suggests itself in the form of the gravitational potential ϕ. So we contemplate the metric

$$ds^2=\phi dt^2+dx^2+dy^2+dz^2. \tag{7}$$

There is a very important aspect of Newton's theory—important both in itself and for our immediate purpose—which we have not yet mentioned. In 1835 Hamilton discovered a variational formulation of Newton's theory, now known as Hamilton's principle, which asserts that the time integral $\int_A^B(T-U)dt$ of the difference of the kinetic energy T and the potential energy U of any mechanical system between two of its states A and B, is stationary. For a test particle traveling between points A and B in a gravitational field ϕ, this becomes

$$\delta\int_A^B(\tfrac{1}{2}v^2-\phi)dt=0, \tag{8}$$

where v is the particle's velocity. (The mass of the particle drops out of the equation, because of the identity $m'=m$.)

Now if the coordinates x,y,z,t of Eq. (7) are the usual coordinates of an inertial frame, then for a moving particle we have

$$v^2=\frac{dx^2+dy^2+dz^2}{dt^2}. \tag{9}$$

With that, Eq. (7) becomes

$$ds^2=(\phi+v^2)dt^2, \tag{10}$$

and the geodesic Eq. (4) reads

$$\delta\int_A^B(\phi+v^2)^{1/2}dt=0. \tag{11}$$

If this were the same as Eq. (8) we would be home! At least, the geodesics of the spacetime with metric (7) would then indeed be the Newtonian paths. But we need not dispair; there are many alternative choices for the metric. Also let us stress that we need not reproduce the Newtonian paths *exactly*. We want to get very close to them—after all, Newton's theory has been exquisitely validated experimentally—but some minute difference could actually be of advantage: it just *might* explain the excessive precession of Mercury!

Apart from not leading to Hamilton's principle for a particle, the tentative metric (7) has another defect: in the complete absence of gravity, i.e., when $\phi\equiv0$, it degenerates into three-dimensionality. Yet we would then wish for a fully four-dimensional though *flat* spacetime in which the geodesics are straight lines, corresponding to constant velocity Newtonian paths. Our next attempt might therefore be

$$ds^2=(V^2-2\phi)dt^2+dx^2+dy^2+dz^2, \tag{12}$$

where V would have to be some universal constant of nature having the dimensions of a velocity, and where we have put -2ϕ rather than ϕ for reasons that will become evident presently. This metric certainly becomes flat while remaining four-dimensional when $\phi\equiv0$. We now show that it also satisfies Hamilton's principle approximately. To get the geodesics of Eq. (12) we simply replace ϕ by $V^2-2\phi$ in Eq. (11)

$$\delta\int_A^B (V^2-2\phi+v^2)^{1/2}dt=0. \tag{13}$$

Now let us suppose that V^2 is very large compared to both v^2 and ϕ. Then, using the binomial expansion, we can transform the integrand of Eq. (13) as follows:

$$\begin{aligned}(V^2-2\phi+v^2)^{1/2}&=V\left(1+\frac{v^2-2\phi}{V^2}\right)^{1/2}\\ &\approx V\left(1+\frac{v^2-2\phi}{2V^2}\right)\\ &=\frac{1}{V}\left(V^2+\frac{1}{2}v^2-\phi\right).\end{aligned} \tag{14}$$

An extra multiplicative constant like $1/V$ in the integrand of a variational principle like Eq. (8) evidently does not affect the solution; nor does an extra additive constant, like V^2 in the last parenthesis of Eq. (14), if t is held fixed at the ends A and B. (That, of course, is consistent with the fact that a potential ϕ is determinate only to within an additive constant.) So the geodesics of the spacetime with Eq. (12) coincide approximately with the Newtonian paths as characterized by Hamilton's principle; and the approximation, stemming from the use of the binomial expansion in Eq. (14), gets better the larger V or the smaller ϕ and v.

Incidentally, we need not apologize for this "tinker's approach" to General Relativity. It is very much in the spirit of what Einstein himself did in the years leading up to 1915, trying one thing after another.[2] Indeed it was he who first and most eloquently stressed the "manmade," model-like character of physical theories.

That said, we can now admit that the metric (12) is still not satisfactory, even though it has the right geodesics. But it violates causality. Consider what it implies in the absence of gravity, when it reduces to

$$ds^2=V^2dt^2+dx^2+dy^2+dz^2. \tag{15}$$

The coordinates are still the x,y,z,t of some inertial frame, indeed of *any* inertial frame, since the argument leading up to Eq. (12) applies in any inertial frame and Eq. (15) merely results from specialization. (It is here that Newtonian relativity—the equivalence of all inertial frames for mechanics—gets encapsuled into the new formalism.) Any two such sets of coordinates satisfying the "4-Euclidean" metric (15) (we can think of Vt as just a fourth Euclidean dimension) are related by four-rotations. Of these, rotations in just Vt and x, as shown in Fig. 6, are special cases. But rotations of this kind play havoc with causality. Consider a particle path AB as shown. In the frame of Vt and x the particle travels from A to B, where A might be the event of its being shot out of a gun and B the event of its shattering a clay pigeon. But in the frame of Vt' and x' it travels from B to A; from the broken pigeon into the smoking barrel of the gun! Considerations of this kind force us, however reluctantly, to contemplate a simple but drastic modification of the metric (12) into

$$\mathbf{ds}^2=(V^2+2\phi)dt^2-dx^2-dy^2-dz^2. \tag{16}$$

This, too, leads to geodesics satisfying Hamilton's principle (8) for large V, and to flat space–time for $\phi\equiv 0$. But additionally it is free of the causality violations of Eq. (12). (The reader must take our word for this; the applicable local coordinate transformations are now essentially the Lorentz transformations and not four-rotations.) But can a "square distance" $\mathbf{ds}^2$ be negative? Well, $\mathbf{ds}^2$ is now not to be regarded as the square of a distance; eventually it is recognized as the square of a displacement vector $\mathbf{ds}$ in spacetime, and the distance ds must be defined as $|\mathbf{ds}^2|^{1/2}$. Spaces with "indefinite" metrics like Eq. (16) (that can be positive, negative, or even zero) are called *pseudo-Riemannian*. Much of the geometry of proper Riemannian spaces applies to them also, especially the theory of geodesics and of curvature. And Special and General Relativity are indeed characterized by metrics of this kind.

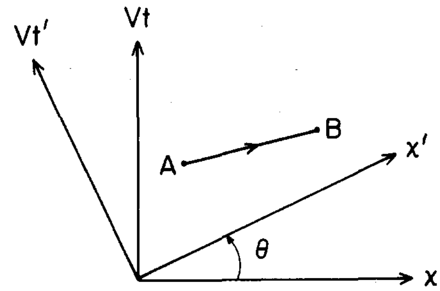

Fig. 6. If space and time in different inertial frames were related by rotations, a signal going from A to B in one frame could go from B to A in another.

VII. SPECIAL RELATIVITY

The metric (16) still has quite a way to evolve before it becomes a viable metric of general-relativistic spacetime. (We shall outline presently how this is done.) However, it has already fulfilled an important purpose in demonstrating the feasibility of modeling Newtonian paths as the geodesics of a "reasonable" spacetime, and thereby "explaining" Galileo's principle. The specialization of Eq. (16) with $\phi\equiv 0$ to gravity-free inertial frames, on the other hand, is already perfect. It is the exact metric of special-relativistic spacetime, provided we write c for the velocity V and identify it with the speed of light—as the modern student will have guessed long ago:

$$\mathbf{ds}^2=c^2dt^2-dx^2-dy^2-dz^2. \tag{17}$$

It may be asked what c has to do with gravity. But the answer is that c has to do with *everything*: it turned out to be part of the structure of space–time, a kind of "conversion factor" from time to space, and almost every relativistic formula contains it. Gravity waves propagate with speed c, black holes of mass m have a radius $2Gm/c^2$, and every mass m has latent energy $E=mc^2$.

Consider a "particle" traveling at speed c in an inertial space–time, i.e., one with a metric of the form (17). Putting $v=c$ in Eq. (9) we find that $\mathbf{ds}^2=0$ for neighboring events at that particle. But since the value of $\mathbf{ds}^2$ (as in every Riemannian or pseudo-Riemannian space) must be independent of the coordinates chosen, it follows that $\mathbf{ds}^2=0$ in all other inertial coordinate systems also, so that the particle has the same velocity c in all inertial frames: c is an *invariant* velocity.

In 1860 *no-one* knew about the invariance of the speed of light (the Michelson–Morley experiment was yet to come) or even about the possibility of an invariant velocity. But since that possibility is made evident by the metric (17), the question *after* accepting Eq. (17) would really have been *which* the invariant velocity is. Moreover, from the comparison of geodesics with Hamilton's principle, one would already know that it must be big. Perhaps, after all, the identification of c with the speed of light might have been made. But independently of that, the rest of Special Relativity could have been derived from the metric (17), at worst with an unknown constant c. As just one example, consider two events at a clock fixed in an inertial system S_0 of coordinates x_0,y_0,z_0,t_0, so that $dx_0=dy_0=dz_0=0$ and $\mathbf{ds}^2=c^2dt_0^2$. In another inertial system S of coordinates x,y,z,t we would then have

$$\mathbf{ds}^2=c^2dt^2\left(1-\frac{dx^2+dy^2+dz^2}{c^2dt^2}\right)=c^2dt^2\left(1-\frac{v^2}{c^2}\right), \tag{18}$$

v being the velocity of S_0 relative to S. This yields the familiar time-dilation formula

$$dt_0=dt(1-v^2/c^2)^{1/2}, \tag{19}$$

which shows that moving clocks go slow. Together with the invariance of c this then yields the familiar length-contraction formula, and so on. Of course, the full Lorentz transformation equations between inertial frames also follow from the invariance of the right-hand side of Eq. (17). But clearly we need not pursue the development of Special Relativity here. On the other hand, we must at least briefly indicate how the approximate metric (16) can be evolved into full General Relativity.

VIII. THE FIELD EQUATIONS OF GENERAL RELATIVITY

The mathematician Riemann might well have had less trouble than did the physicist Einstein with the development of the metric (16). The task is twofold: first, to eliminate all direct dependence on Newtonian theory (as in the use of the Newtonian potential ϕ and of Newtonian inertial frames), and then to make spacetime theory truly geometric in the spirit of Gauss, namely to express it in a form independent of any special choice of the coordinates x_1,x_2,x_3,x_4 for labeling events.

One must therefore contemplate a general pseudo-Riemannian four-dimensional metric of the form (6), with signature $+---$. (This refers to the distribution of signs when the metric is expressed as a sum of squared differentials at any given point by a linear transformation of coordinates.) Such a general metric contains 10 as yet undetermined coefficients g_{ij}, which, in view of Eq. (16), are regarded as the *relativistic potentials*. Just as in Newtonian theory (which has one potential ϕ) and in Maxwell's theory (which has four potentials ϕ, A_1,A_2,A_3), the first derivatives of the potentials (the g_{ij}) determine the (now geodesic) equations of motion of the theory. By analogy with Poisson's differential formulation of Newton's inverse-square law, $\nabla^2\phi=4\pi G\rho$, and with the corresponding Maxwell equations, we might expect that out of the *second* derivatives of the g_{ij} one can construct the *field equations* which link the field g_{ij} to the sources. And this is precisely what Einstein succeeded in doing. Furthermore, his field equations reduce to Poisson's equation in the case of weak fields, just as the geodesic equations reduce to Hamilton's principle in the case of weak fields and slow motions. (In the context of relativity, the field of the sun, for example, is exceedingly weak, i.e., $\phi \ll c^2$, and the motions of the planets are exceedingly slow, i.e., $v \ll c$.)

Riemann might not have been able to anticipate Einstein's *full* field equations, which are based on a special-relativistic analysis of the 10 parameters characterizing a continuous distribution of sources (the components of the so-called momentum–stress–energy tensor) and which incorporate Einstein's discovery $E=mc^2$. But he could rather easily have anticipated the *vacuum* field equations governing source-free regions like the exterior field of the sun. He would simply have had to "contract" the very tensor associated with his name, the Riemann curvature tensor, and set the result equal to zero. A certain analogy between, on the one hand, curvature having to do with the spread of geodesics, and, on the other hand, a spread of geodesics having to do with tidal forces in Newtonian theory, could actually have suggested this approach.[3] And the exact precession of the orbit of Mercury would then have been an immediate consequence, ripe for the picking.

Two features, in particular, of General Relativity set it apart from theories such as Newton's or Maxwell's. The first is that the field equations are nonlinear in the potentials, and the second is that the field *is* the geometry. The main implication of nonlinearity is that solutions cannot be superimposed. If we simply add the fields of the parts of a star in Newton's theory, we get the field of the star. Not so in Einstein's theory, and we should not expect it: here the negative gravitational binding energy holding the various parts together has itself a mass (via $E=mc^2$) and this reduces the sum of the fields of the parts. The dilemma inherent in the coincidence of field and geometry is that, on the one hand, we cannot describe the sources properly unless we know the geometry in which they are embedded, and, on the other hand, we do not know the geometry until we have solved the field equations which, of course, involve the sources. This forces us to work "from both ends" at the same time—quite a new technique.

IX. CONCLUDING REMARKS

General Relativity, apart from reproducing all the results of Newtonian theory to within the classical accuracy of observation, additionally yields the correct orbit of Mercury and also takes care of the other three difficulties of Newtonian theory that we mentioned: the Galilean principle is explained by the geodesic law of motion, the absolute inertial frames have been replaced by a spacetime susceptible to the influence of matter via the field equations, and last, General Relativity provides a completely satisfactory dynamics and geometry for the universe as a whole. Beyond that, of course, like any good theory, it has led to many new and unexpected predictions, such as the existence of gravity waves and of black holes, to mention just the two best known.

Furthermore, General Relativity linked the propagation of light to gravity. One can prove that light follows *null* geodesics in spacetime, i.e., geodesics having $\mathbf{ds}^2=0$ all along them. Since the geodesics are determined by the metric, which in turn is determined by the sources, we see that light paths are also dependent on the latter: just like particle paths, light paths are bent in the presence of gravitating sources. Of course, in Newton's theory, too, particles (photons) traveling at the speed of light are deflected (as are particles traveling at any other speed.) But it turns out that the general-relativistic orbits of photons (unlike the orbits of slow particles) can differ radically from their Newtonian counterparts. For example, radio signals from distant quasars passing the edge of the sun on their way to us should be bent by *twice* the Newtonian value according to General Relativity. And this is confirmed by observation.

In any presentation of General Relativity, however sketchy, the opportunity should not be missed to drive home to the student an important insight on the nature of physical theories. Ever since Fancis Bacon, it had been believed that the laws of Nature were there to be "discovered," if only one made the right experiments. Einstein taught us differently. He stressed the vital role of human inventiveness in the process. Newton "invented" the force of gravity to explain the motion of the planets. Einstein "invented" curved space–time and the geodesic law; in his theory there is *no* force of gravity. If two such utterly different mathematical models can (almost) both describe the same observations, surely it must be admitted that physical theories do not tell us what nature *is*, only what it *is like*. The marvel is that nature seems to go along with some of the "simplest" models that can be constructed in the context of various mathematical formalisms.

Finally, as a kind of summary, the student could well be left with a modern definition and characterization of relativity. Thus, relativity should be regarded *primarily* as a new theory of space and time, in which these two concepts meld into a pseudo-Riemannian spacetime; *secondarily* relativity is the theory of a new physics consistent with this new space–time background. Special Relativity deals with physics in *flat* spacetime, which ideally exists only in the total absence of gravitating matter. For, according to the field equations of General Relativity, the presence of matter curves the spacetime. General Relativity is thus the modern theory of gravity. But it also deals with the rest of physics in curved spacetime. In practice this can often be avoided by treating sufficiently small portions of curved spacetime as flat, and simply using Special Relativity. Such an approach is analogous to treating a small portion of a curved surface (e.g., our backyard) as flat, and applying plane Euclidean geometry to it.

Perhaps to a bright student we have raised more questions than we have answered. But if this whets the appetite for a deeper study of relativity, then our purpose will have been amply served.[4]

[1]B. Riemann, "Über die Hypothesen, welche der Geometrie zugrunde liegen," Inaugural Lecture (1854). Posthumously published in Nachr. Ges. Wiss. Göttingen, **13**, 133–153 (1868). For an annotated extract from this lecture, in English, see C. W. Kilmister, *General Theory of Relativity* (Pergamon, Oxford, 1973), pp. 101–122.

[2]Good accounts of Einstein's progression towards General Relativity can be found in Einstein's own "Autobiographical notes" in *Albert Einstein: Philosopher–Scientist*, edited by P. A. Schilpp (The Library of Living Philosophers, Evanston, IL, 1949), pp. 3–95, but especially pp. 63–81; also in E. Whittaker, *A History of the Theories of Aether and Electricity* (T Nelson, London, 1953), Vol. II, Chap. V; and especially in A. Pais, *Subtle is the Lord···—The Life and Science of Albert Einstein* (Clarendon, Oxford, 1982), Chaps. 9–14.

[3]See, for example, W. Rindler, *Essential Relativity*, rev. 2nd ed. (Springer, New York, 1979), pp. 133–136.

[4]Some books for further reading might be the following: G. F. R. Ellis and R. M. Williams, *Flat and Curved Space-Times* (Clarendon, Oxford, 1988); R. D'Inverno, *Introducing Einstein's Relativity* (Clarendon, Oxford, 1992); I. R. Kenyon, *General Relativity* (Oxford University Press, Oxford, 1990); B. F. Schutz, *A First Course in General Relativity* (Cambridge University Press, Cambridge, 1985); W. Rindler, in Ref. 3, especially Chaps. 1 and 7.

What is truth? A course in science and religion

Peter J. Brancazio
Professor of Physics, Brooklyn College, CUNY, Brooklyn, New York 11210

(Received 12 November 1993; accepted 12 May 1994)

Are science and religion independent or conflicting systems of thought? This question was the focus of a course on science and religion given at Brooklyn College. The class explored the underlying metaphysical assumptions, sources of knowledge, and methodologies of science and religion from philosophical and historical perspectives. The conclusion is that science and religion are basically independent modes of inquiry, but they have come into conflict over questions about the origin, history, and nature of the physical world. These are areas in which science has been far more successful in providing fruitful explanations and predictions. Nevertheless, the limitations of science and the scientific method must be understood and respected.

The subject of the relationship between science and religion never fails to stimulate passionate debate. The classic text on the subject, written in 1895, referred to their interactions as "warfare." [1] Scientists have come in conflict with religious authorities dating back to ancient Greece; the Galileo affair still resonates after three and a half centuries. In our times, the conflict between creationism and evolution reminds us that a full accommodation has yet to be reached. Most recently, the publication in the *American Journal of Physics* of the review of a book on science and religion stimulated one of the more spirited and sustained exchange of letters to the editor in recent memory.[2]

The purpose of this article is to report on a course on Science and Religion given at Brooklyn College in the Fall

