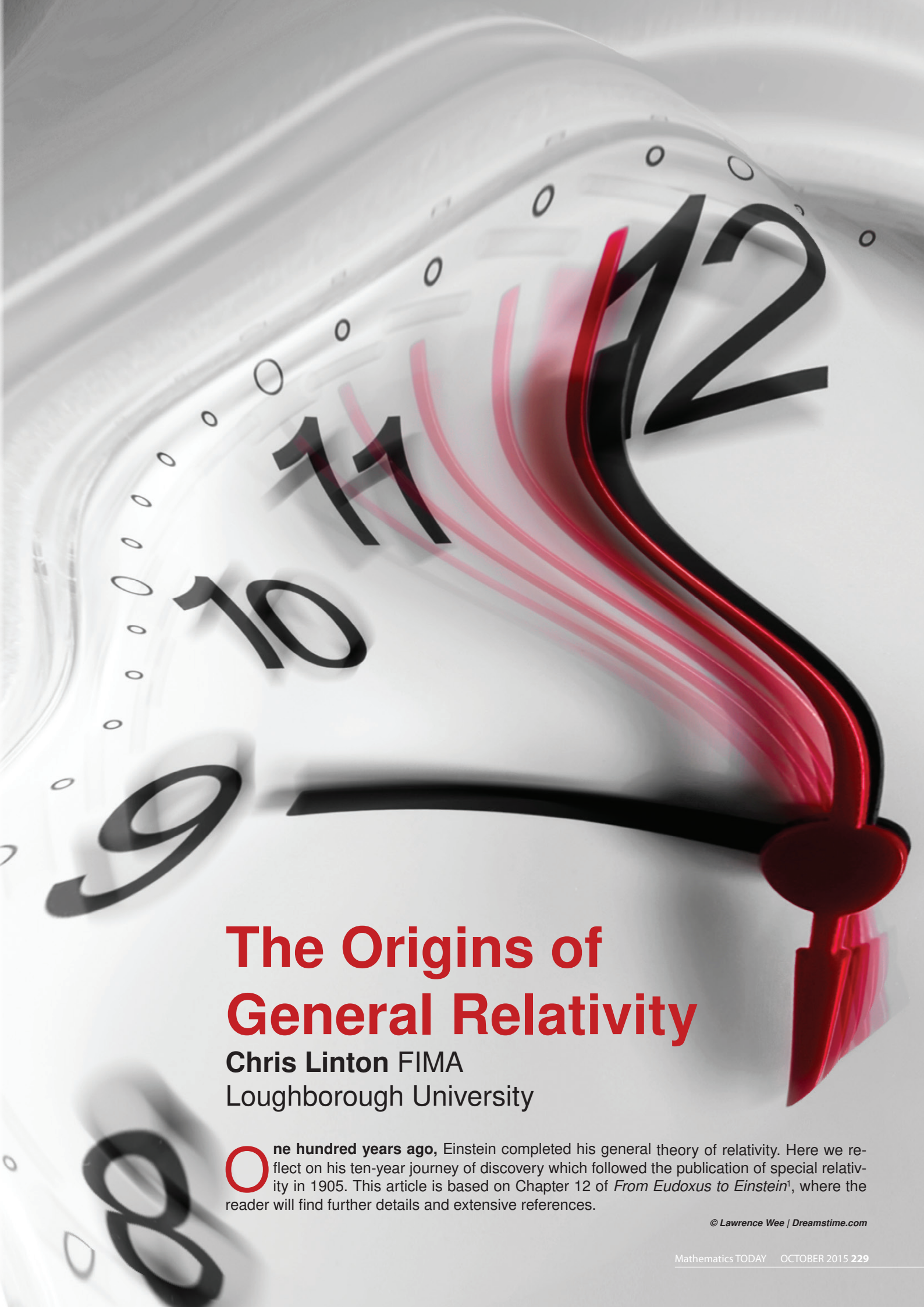




The Origins of General Relativity

Chris Linton FIMA
Loughborough University

One hundred years ago, Einstein completed his general theory of relativity. Here we reflect on his ten-year journey of discovery which followed the publication of special relativity in 1905. This article is based on Chapter 12 of *From Eudoxus to Einstein*¹, where the reader will find further details and extensive references.

© Lawrence Wee / Dreamstime.com

In the latter half of the 19th century, tables used to determine Mercury's position were notoriously inaccurate, in contrast to those for all the other planets. Mercury's orbit can be thought of as an ellipse which slowly rotates so that the point of closest approach, the perihelion, slowly precesses around the sun. Much of this perihelion advance is due to the perturbative effect of the other planets, notably Venus and Jupiter, and calculations showed that these effects caused an advance of roughly $530''$ per century (about $1\frac{1}{4}''$ per orbit) leaving $43''$ per century of the observed value unexplained.

Many alternatives were suggested as possible explanations, including the possibility of an additional planet inside Mercury's orbit or modifications to the inverse square law of gravitation, but nothing particularly satisfactory emerged. So when Einstein was reflecting on how special relativity might be extended to encompass a new theory of gravitation, the problem of Mercury's perihelion was firmly in his mind. In a letter in 1907 he wrote that he was 'busy on a relativistic theory of the gravitational law with which I hope to account for the still unexplained secular change of the perihelion motion of Mercury' but it wasn't until 1911 that Einstein's serious attack on a new theory of gravitation began.

One stimulus for this new activity was Einstein's realisation that the bending of light might be capable of experimental verification by considering rays of light passing close to the sun during a solar eclipse. Since (as a consequence of special relativity) inertial mass depends on energy through the relation $E = mc^2$, the equivalence of inertial and gravitational mass implies that the same must be true for the latter. This leads to what Einstein described as a 'consequence which is of fundamental importance', namely that the speed of light is a function of the gravitation potential Φ and is given, to the first approximation, by

$$c = c_0(1 + \Phi/c_0^2), \quad (1)$$

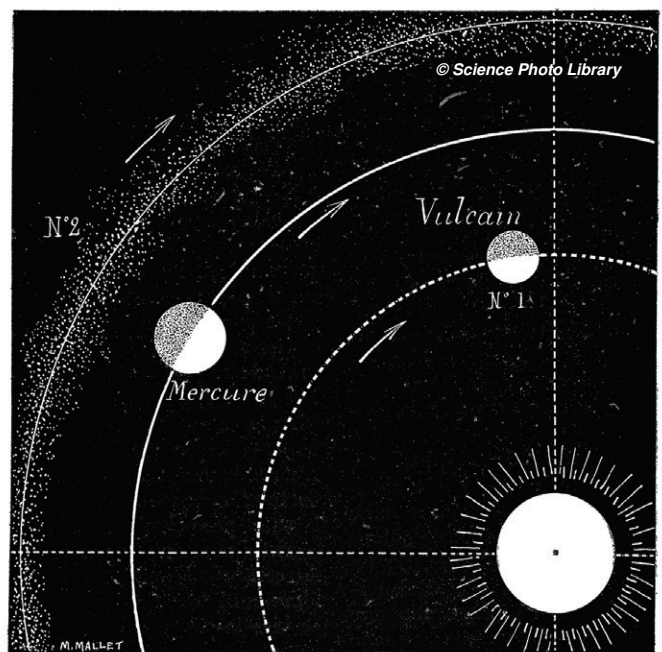
where c_0 is the speed of light from special relativity. Thus in order to extend special relativity to accommodate the equivalence of inertial and gravitational mass, Einstein had to modify one of the fundamental postulates of the original theory. Reconciling special relativity with gravity was not going to be easy.

Since the speed of light varies in a gravitational field, it follows from Huygens' principle that it will be bent, and Einstein calculated the magnitude of the deflection for a light ray coming from infinity and passing at a distance Δ from a point mass M , as $2GM/c^2\Delta$, where G is the universal gravitational constant. For the sun, Einstein computed the value $0''.83$ and remarked that 'it would be a most desirable thing if astronomers would take up the question here raised'. Einstein (just like the rest of the physics community) was entirely unaware that almost exactly the same prediction had been made over 100 years before him by combining Newtonian optics and mechanics! If we envisage light as being made up of particles, then these will be attracted by gravity just like any other matter. It is straightforward to compute the deflection of a particle in a hyperbolic orbit passing close to a point mass M given that its speed at infinity is c and this appears to have been first done in 1784 by Cavendish, who obtained the value $2 \sin^{-1}(\epsilon/(1+\epsilon))$, where $\epsilon = GM/c^2\Delta$. This agrees with Einstein's formula, since ϵ is small. A very similar calculation was carried out in 1801 by von Soldner, with the same conclusion. Cavendish never published his result, but von Soldner's calculation appeared in a major astronomical journal – and was largely ignored.

In his 1911 paper, Einstein also quantified the gravitational red shift, showing that, again to the first approximation, the spectral lines when measured on the earth should be shifted towards the red by a frequency of Φ/c^2 , Φ being the difference in potential between the surface of the sun and the earth. Einstein calculated this figure as 2×10^{-6} and noted that such an effect had been observed but attributed to physical phenomena on the surface of the sun. The measurement of the gravitational red shift is in fact extremely difficult as it is masked by convection currents in the solar atmosphere, which lead to Doppler shifts of the same order of magnitude.

At this point, Einstein had gone about as far as possible toward a new theory of gravity within the kinematical framework of special relativity. But quite what new idea was required to crack the problem he didn't know. He wasn't the only person working on gravitation at this time. In 1912 Abraham, a staunch opponent of relativity, published a theory in which the direction of the force of gravity is given by ∇c , with the speed of light given as a function of the gravitational potential by $c = c_0(1 + 2\Phi/c_0^2)^{1/2}$, which agrees with Einstein's (1) to first order in $1/c_0^2$. Since $E = mc^2$ and the total energy of a body in a gravitational field is a function of the potential Φ , it follows that either m or c (or both) must depend on Φ . Both Einstein and Abraham had $c = c(\Phi)$, but theories in which $m = m(\Phi)$ and c retained its constant special relativity value were also put forward.

In all these attempts to extend relativity to include gravitation the field was determined from a single scalar function (as it is in Newtonian mechanics) and Einstein developed his own scalar theory of a static gravitational field during the same period. Although the theory had problems, one positive outcome was the unpleasant realisation that the equations of gravitation had to be nonlinear since the gravitational field possesses energy which acts as its own source. This then led Einstein to the reluctant conclusion that the principle of equivalence could only hold locally. In other words it is not possible, in general, to find a global coordinate transformation which removes the effect of gravity completely, but that this is possible at each point in space-time.



Artwork from the French popular science weekly *La Science Illustrée* (1895), illustrating Le Verrier's theory that an undiscovered planet (Vulcan) explained the precession of Mercury.

It was in 1912 that Einstein was struck by the similarities between the problems he was facing and the theory of surfaces due to Gauss. With the help of his friend Marcel Grossman he learned about the ‘absolute differential calculus’ developed through the work of Riemann, Christoffel, Ricci and Levi-Civita, and began to investigate whether gravitation could be described through a curved space defined via the metric

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (2)$$

where $g_{\mu\nu}$ (which depends on the coordinates x^μ) is a symmetric tensor of rank 2. Here we have denoted x, y, z and ct , by x^1, x^2, x^3 and x^4 , respectively, and used a notational convenience which Einstein himself introduced in 1916 and which is now referred to as the Einstein summation convention. If the index in an expression occurs twice, once as a subscript and once as a superscript (μ and ν in the above equation), then a summation is implied. Hence (2) is shorthand for $ds^2 = \sum_{\mu=1}^4 \sum_{\nu=1}^4 g_{\mu\nu} dx^\mu dx^\nu$. The difficulties that Einstein faced when developing general relativity were not helped by the fact that his notation was initially much more cumbersome.

This would be a radical departure from any previous type of theory, in which the underlying structure of space and time was given to begin with. In his new approach, the geometry of space–time was one of the things which gravity had to explain. It also made everything much more complicated since the single variable c had been replaced by the ten unknowns $g_{\mu\nu}$. Einstein clearly found the work incredibly difficult. In 1912 he wrote to Sommerfeld:

At present I occupy myself exclusively with the problem of gravitation and now believe that I shall master all difficulties with the help of a friendly mathematician here. But one thing is certain, in all my life I have labored not nearly as hard, and I have become imbued with great respect for mathematics, the subtler part of which I had in my simple-mindedness regarded as pure luxury until now. Compared with this problem, the original relativity is child’s play.

The collaboration with Grossman led to the publication in 1913 of what is often referred to as the *Entwurf* theory. This was Einstein’s first attempt at a tensor theory and it came remarkably close to the final theory that would emerge just over two years later.

It was Minkowski who had first formulated the equations of special relativity in modern tensor form, but Einstein was initially unimpressed by this formal simplification to the theory (he described it as ‘superfluous learnedness’). The Minkowski metric can be written

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (3)$$

where $\eta_{\mu\nu} = 0$ if $\mu \neq \nu$ and $\eta_{\mu\mu} = 1, 1, 1$ and -1 , for $\mu = 1, 2, 3$ and 4 , respectively. The principle of equivalence as understood by Einstein in 1913 is equivalent to the condition that at any point in space–time there is a coordinate transformation which reduces the metric tensor $g_{\mu\nu}$ to the Minkowski tensor $\eta_{\mu\nu}$. Locally, space–time has the structure of special relativity.

In the *Entwurf* theory, the motion of a particle in free fall within a gravitational field given by (2) is determined from the variational principle $\delta \int ds = 0$. In other words, in the absence of any forces, objects move between two points along the shortest path just as they do in special relativity, but now length is being

measured by the line element given in (2) and so the resulting trajectories are not straight lines. Gravity is thus no longer a force but instead is considered as a deformation of Minkowski’s space–time. A helpful analogy is to think of gravity as a curved surface which can be approximated locally by the tangent plane to the surface. To complete the theory, equations which relate the metric $g_{\mu\nu}$ to the distribution of matter are needed – these are the so-called field equations of gravitation – and this is where Einstein and Grossman came up short.

The core of general relativity as we know it today is that all frames of reference should be treated equally, none are preferred. However, the field equations in the *Entwurf* paper are not generally covariant, but covariant only with regard to linear transformations, and Einstein thought the theory marred by an ‘ugly dark spot’. Although he defended it in public, Einstein was never fully convinced by the *Entwurf* theory and it was not received well by other physicists.

Einstein and Grossman did not comment on the effect of their theory on Mercury’s perihelion, though Einstein did subsequently work it out, in collaboration with an old friend from his student days Michele Besso. Their value of $18''$ per century was not published, but was confirmed by Droste in 1915. In fact, throughout the development of general relativity, Einstein rarely mentioned perihelion shifts, but paid much more attention to light deflection and red-shift calculations. This was probably because, since the effects of other matter may well be important, he was not confident that a new theory of gravitation would eventually yield all of the missing $43''$. The same was true of the other competitors in the race for a new theory of gravity. However, immediately after Einstein arrived at the generally covariant field equations for gravity in 1915, he wrote to Sommerfeld and gave three reasons why he had turned his back on the *Entwurf* theory, one of which was the fact that it did not give the correct advance for the perihelion of Mercury.

One thing that was missing in the Einstein–Grossman theory, but which Einstein derived in 1914, was the relativistic equation of free fall (the geodesic equation), though this was well established in the theory of Riemannian geometry. The principle of equivalence implies that in such a situation one can find local coordinates ξ^α such that

$$\frac{d^2 \xi^\alpha}{ds^2} = 0$$

with ds^2 given by its special relativistic form (3). Now consider another coordinate system x^μ . We have, using Einstein’s summation convention,

$$\frac{d^2 \xi^\alpha}{ds^2} = \frac{\partial}{\partial x^\nu} \left(\frac{\partial \xi^\alpha}{\partial x^\mu} \frac{dx^\mu}{ds} \right) \frac{dx^\nu}{ds}.$$

If we evaluate the derivative and multiply by $\partial x^\lambda / \partial \xi^\alpha$ we obtain the geodesic equations

$$\frac{d^2 x^\lambda}{ds^2} + \Gamma_{\mu\nu}^\lambda \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0, \quad (4)$$

where

$$\Gamma_{\mu\nu}^\lambda = \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \quad (5)$$

is the Christoffel symbol of the second kind, also called the affine connection, or simply the connection coefficient. Applying the same transformation to the line element ds , which must be invariant under a coordinate transformation since it is a property

of the underlying space–time, we find that $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$, where the metric tensor takes the form

$$g_{\mu\nu} = \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta}.$$

Within the framework of relativity there are no forces acting on an object in free fall and so the effect of gravity is determined by the affine connection. By itself this is not particularly useful, since (5) requires knowledge of the local inertial coordinates at each point. However, it can be shown by direct calculation that

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\sigma} \left(\frac{\partial g_{\nu\sigma}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right)$$

and hence the metric tensor can be considered as the gravitational potential.

The final twists and turns in the creation of general relativity are evident from a series of four dramatic communications made by Einstein to the Prussian Academy during November 1915. The first was on the 4th and the second a week later. In these Einstein began the return towards general covariance which had been his goal right from the start, but he was still held back by some of his previous misconceptions. It was Mercury that provided the key. A calculation of the perihelion motion based on the equations he was now working with yielded 43'' per century for Mercury's advance. Overjoyed, he wrote that his theory

explains ... quantitatively ... the secular rotation of the orbit of Mercury, discovered by Le Verrier, ... without the need of any special hypothesis.

In his communication of this result on the 18th he also noted that his new theory gave a deflection of light which was twice his earlier prediction. There were still some problems, but the quantitative success with Mercury meant there was no doubt that he was on the right track. In the course of his analysis, Einstein crucially realised that the form of the equations in the case of a weak, static field could be more general than he had hitherto supposed. With the stumbling block removed, Einstein soon grasped the true nature of the solution he was looking for and on the 25th he communicated the final form for the field equations, concluding that 'the general theory of relativity is closed as a logical structure'. The change to the final set of equations did not affect the calculations of perihelion shift and light deflection of the week before.

In Newtonian theory the gravitational potential Φ satisfies Poisson's equation $\nabla^2 \Phi = -4\pi\rho G$, where ρ is the mass density of the source of the gravitational field. In general relativity gravity is determined from the tensor $g_{\mu\nu}$ and so, if we want any chance of recovering the Newtonian formula in an appropriate limit, the role of the mass density must also be played by a second-order tensor. This is the energy–momentum stress tensor $T_{\mu\nu}$ first introduced by Minkowski, which serves to define the distribution of energy and momentum throughout space–time. Where the energy–momentum tensor is zero, general relativity must reduce to special relativity and hence whatever is to appear in place of $\nabla^2 \Phi$ must in such cases be zero. This provides a vital clue.

The most general tensor that can be constructed from the metric tensor and its first and second derivatives, linear in the second derivatives, is the Riemann–Christoffel curvature tensor $R_{\mu\nu}^\lambda$, defined by

$$R_{\mu\kappa\nu}^\lambda = \frac{\partial \Gamma_{\mu\kappa}^\lambda}{\partial x^\nu} - \frac{\partial \Gamma_{\mu\nu}^\lambda}{\partial x^\kappa} + \Gamma_{\mu\kappa}^\eta \Gamma_{\nu\eta}^\lambda - \Gamma_{\mu\nu}^\eta \Gamma_{\kappa\eta}^\lambda.$$

To obtain a tensor of rank two, we can form the contraction $R_{\mu\nu} = R_{\mu\lambda\nu}^\lambda$ which is known as the Ricci tensor and we can also make use of $R = g^{\mu\nu} R_{\mu\nu}$, which is the curvature scalar.

The crucial property of the curvature tensor for our discussion here is that if we are in the situation of special relativity, where the metric tensor is $\eta_{\mu\nu}$, then $R_{\mu\kappa\nu}^\lambda = R_{\mu\nu} = R = 0$. On 11 November, Einstein proposed that the field equations of general relativity are

$$R_{\mu\nu} = -\kappa T_{\mu\nu},$$

where κ is a constant. This then reduces to

$$R_{\mu\nu} = 0 \tag{6}$$

in empty space. Einstein had been here before, but he and Grossman had earlier been led off the scent with their erroneous belief that (6) did not have the correct form in the Newtonian limit.

When Einstein managed to show that (6) led to exactly the right advance for Mercury's perihelion he knew that he was very close, and it did not take him long to appreciate that no additional assumptions were required if he wrote $R_{\mu\nu} = -\kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T)$, where $T = T_{\mu}^{\mu}$, or equivalently,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\kappa T_{\mu\nu}. \tag{7}$$

These are the field equations of general relativity from which the metric tensor is to be determined. An analysis of the Newtonian limit reveals that the coupling constant κ must have the value $8\pi G/c^4$. Note that in empty space we still have (6).

Let us turn now to the motion of Mercury. This is essentially a one-body problem as we can neglect the effects of Mercury's mass on the gravitational field of the sun. Rather than following Einstein's calculation, we will begin with the exact solution to Einstein's field equations found in 1916 by Schwarzschild. If we assume that the sun's field is static and isotropic then we obtain from (6), with coordinates (r, θ, ϕ) which far from the origin are the usual spherical polar coordinates,

$$ds^2 = \frac{dr^2}{1 - \alpha/r} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) - \left(1 - \frac{\alpha}{r}\right) c^2 dt^2. \tag{8}$$

Here $\alpha = 2GM/c^2$, where M is the mass of the sun.

A lengthy calculation shows that the geodesics for the metric (8) in the case $\theta = \frac{1}{2}\pi$ (corresponding to the orbital plane of Mercury) are the solutions to the equation

$$\frac{d^2 u}{d\phi^2} + u = \frac{GM}{h^2} + 3GM \frac{u^2}{c^2},$$

in which $u = 1/r$ and $h = r^2 \dot{\phi}$. This is the same as the equivalent Newtonian equation except for the presence of the final relativistic term and has as a first integral

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{2GM}{h^2}u + \frac{2GM}{c^2}u^3 + E, \tag{9}$$

where E is a constant of integration. At aphelion and perihelion $du/d\phi = 0$ and then

$$\alpha u^3 - u^2 + \beta u + E = 0,$$

where $\alpha u = 2GM/c^2 r \ll 1$ and $\beta = 2GM/h^2$. This is a cubic equation, two of whose roots u_1 and u_2 , say, are the reciprocal distances at aphelion and perihelion, respectively. Given that the

sum of the roots must be $1/\alpha$, the third root u_3 must be large. It follows from (9) that

$$\frac{du}{d\phi} = [\alpha(u - u_1)(u_2 - u)(u_3 - u)]^{1/2}.$$

If we replace u_3 by $1/\alpha - u_1 - u_2$ we find that, to first order in α ,

$$\frac{d\phi}{du} = \frac{1 + \frac{3}{2}\alpha v_+ + \frac{1}{2}\alpha(u - v_+)}{[v_-^2 - (u - v_+)^2]^{1/2}},$$

where $v_{\pm} = \frac{1}{2}(u_2 \pm u_1)$, and the angle between aphelion and the next perihelion is

$$\int_{u_1}^{u_2} \frac{d\phi}{du} du = \pi \left(1 + \frac{3}{2}\alpha v_+\right).$$

If we substitute the Newtonian values $r_1 = a(1 + e)$ and $r_2 = a(1 - e)$, where a and e are the semi-major axis and eccentricity of Mercury's orbit, respectively, we find that the advance in the perihelion per revolution is

$$\frac{6\pi GM}{c^2 a(1 - e^2)}.$$

For the sun, $GM/c^2 \approx 1475$ m and for Mercury $a(1 - e^2) \approx 55.46 \times 10^9$ m so that the advance comes out to be $0''.1034$ per revolution, or a little under $43''$ per century. Stunning as this confirmation of observation might appear, it was going to take more than one experimental success for people to willingly accept a theory which reinterprets the whole fabric of space and time. Moreover, there are a number of other possible influences on Mercury's motion which are hard to quantify. The advance in technology in the second half of the 20th century has led to numerous new tests for general relativity and the theory has passed each one (with many competing theories failing) and so the grounds for accepting Einstein's theory are now extremely strong. But this was not the case when it was published.

In 1916, in his first full-scale exposition of general relativity, Einstein wrote

These equations, which proceed by the method of pure mathematics, from the requirement of the general theory of relativity, give us ... to a first approximation Newton's law of attraction, and to a second approximation the explanation of the motion of the perihelion of the planet Mercury discovered by Le Verrier... These facts must, in my opinion, be taken as convincing proof of the correctness of the theory.

Others disagreed. For example, von Laue objected on the grounds that the calculations assumed that the sun and Mercury were point masses, and unlike in the Newtonian theory there was no reason to believe that one could treat extended bodies this way. Einstein recognised that more experimental confirmation should be sought and at the end of the paper we find his calculations for the gravitational red shift, the bending of light (which is now double his original 1911 value), and the motion of Mercury.

As mentioned above, red-shift observations were insufficiently precise to be of much use. On the other hand, observations of light deflection made in 1919 made Einstein an international superstar. Given the negative reception that special relativity had in England and the fact that the First World War had made

German journals virtually inaccessible to British scientists, it is perhaps surprising that it was a British team which provided the first new empirical justification for Einstein's theory. Einstein had at least one significant champion in England, the secretary of the Royal Astronomical Society Arthur Eddington, and he persuaded Dyson, the Astronomer Royal, to support two expeditions to measure the deflection of light during the solar eclipse of May 1919. One team set out to Sobral in Brazil, and another (led by Eddington) travelled to Principe, off the west coast of Africa. The results from these expeditions were far from conclusive. The Sobral group reported a deflection of $1''.98 \pm 0''.16$ and those in Principe found $1''.61 \pm 0''.40$, compared with the general relativistic value of $1''.75$, but Dyson proudly announced that Einstein's prediction had been confirmed. On 7 November 1919 the London *Times* carried an article headed with 'Revolution in Science/New Theory of the Universe/Newtonian Ideas Overthrown' and the *New York Times* carried some equally dramatic headlines two days later.

General relativity is a highly technical mathematical theory, yet it is firmly grounded on experiment. The guiding principles behind the theory – that all frames of reference are to be treated equally, special relativity should be recovered in the absence of gravitation, and Newtonian mechanics are appropriate when gravitational fields are weak – all have a sound empirical basis. Moreover, as more experimental tests have been performed, confidence in general relativity has grown. But the non-linear nature of the theory makes it seriously challenging and there is still plenty more to discover lying hidden within the deceptively simple looking equation (7).

Notes

1. This article has drawn upon material from within C.M. Linton, 'Mercury and Relativity', in C.M. Linton, *From Eudoxus to Einstein, A History of Mathematical Astronomy*, (2004) © C.M. Linton 2004, published by Cambridge University Press, reproduced with permission.

