EINSTEIN’S THEORY OF RELATIVITY

(February 3, 1929)

presented to the general public in terms that the average person could understand (or so he thought).

THE HISTORY OF FIELD THEORY

(‘Old and New developments of Field Theory’)

By Albert Einstein

While physics wandered exclusively in the paths prepared by Newton, the following conception of physical reality prevailed: matter is real, and matter undergoes only those changes which we conceive as movements in space. Motion, space and also time are real forms. Every attempt to deny the physical reality of space collapses in face of the law of inertia. For if acceleration is to be taken as real, then that space must also be real within which bodies are conceived as accelerated.

Newton saw this with perfect clarity and consequently he called space "absolute". In his theoretical system, there was a third constituent of independent reality; the motive force acting between material particles, such forces being considered to depend only on the position of the particles. These forces between particles were regarded as unconditionally associated with the particles themselves and as distributed spatially according to an unchanging law.

The physicists of the nineteenth century considered that there existed two kinds of such matter, namely, ponderable matter and electricity. The particles of ponderable matter were supposed to act on each other by gravitational forces under Newton's law, the particles of electrical matter by Coulomb forces also inversely proportional to the square of the distance. No definite views prevailed regarding the nature of the forces acting between ponderable and electrical particles.

THE OLD THEORY OF SPACE

Mere empty space was not admitted as a carrier for physical changes and processes. It was only, one might say, the stage on which the drama of material happenings was played. Consequently Newton dealt with the fact that light is propagated in empty space by making the hypothesis that light also consists of material particles interacting with ponderable matter through special forces. To this extent Newton's view of nature involved a third type of material particle, though this certainly had to have very different properties from the particles of the other forms of matter. Light particles had, in fact, to be capable of being formed and of disappearing. Moreover, even in the eighteenth century it was already clear from experience that light travelled in empty space with a definite velocity, a fact which obviously fitted badly into Newton's theoretical system, for why on earth should the light particles not be able to move through space with any arbitrary velocity?

It need not, therefore, surprise us that this theoretical system, built up by Newton with his powerful and logical intellect, should have been overthrown precisely by a theory of light. This was brought about by the Huygens-Young-Fresnel wave theory of light which the facts of interference and diffraction forced on stubbornly resisting physicists. The great range of phenomena, which could be calculated and predicted to the finest detail by using this theory, delighted physicists and filled many fat and learned books. No wonder then that the learned men failed to notice the crack which this theory made in the statue of their eternal goddess. For, in fact, this theory upset the view that everything real can be conceived as the motion of particles in space. Light waves, were, after all, nothing more than undulatory states of empty space, and space thus gave up its passive role as a mere stage for physical events. The other hypothesis patched up the crack and made it invisible.

The ether was invented, penetrating everything, filling the whole of space, and was admitted as a new kind of matter. Thus it was overlooked that by this procedure space itself had been brought to life. It is clear that this had really happened, since the ether was considered to be a sort of matter which could nowhere be removed. It was thus to some degree identical with space itself; that is, something necessarily given with space. Light was thus viewed as a dynamical process undergone, as it were by space itself. In this way the field theory was born as an illegitimate child of Newtonian physics, though it was cleverly passed off a first as legitimate.

To become fully conscious of this change in outlook was a task for a highly original mind whose insight could go straight to essentials, a mind that never got stuck in formulas. Faraday was this favoured spirit. His instinct revolted at the idea of forces acting directly at a distance which seemed contrary to every elementary observation. If one electrified body attracts or repels a second body, this was for him brought about not by a direct action from the first body on the second, but through an intermediary action. The first body brings the space immediately around it into a certain condition which spreads itself into more distant parts of space, according to a certain spatio-temporal law of propagation. This condition of space was called "the electric field." The second body experiences a force because it lies in the field of the first, and vice versa. The "field" thus provided a conceptual apparatus which rendered unnecessary the idea of action at a distance. Faraday also had the bold idea that under appropriate circumstances fields might detach themselves from the bodies producing them and speed away through space as free fields: this was his interpretation of light.

Maxwell then discovered the wonderful group of formulae which seems so simple to us nowadays and which finally build the bridge between the theory of electromagnetism and the theory of light. It appeared that light consists of rapidly oscillating electromagnetic fields.

After Hertz, in the '80s of the last century, had confirmed the existence of the electromagnetic waves and displayed their identity with light by means of his wonderful experiments, the great intellectual revolution in physics gradually became complete. People slowly accustomed themselves to the idea that the physical states of space itself were the final physical reality, especially after Lorentz had shown in his penetrating theoretical researches that even inside ponderable bodies the electromagnetic fields are not to be regarded as states of the matter, but essentially as states of the empty space in which the material atoms are to be considered as loosely distributed.

DISSATISFIED WITH DUAL THEORY

At the turn of the century physicists began to be dissatisfied with the dualism of a theory admitting two kinds of fundamental physical reality: on the one hand the field and on the other hand the material particles. It is only natural that attempts were made to represent the material particles as structures in the field, that is, as places where the fields were exceptionally concentrated. Any such representation of particles on the basis of the field theory would have been a great achievement, but in spite of all efforts of science it has not been accomplished. It must even be admitted that this dualism is today sharper and more troublesome that it was ten years ago. This fact is connected with the latest impetus to developments in quantum theory, where the theory of the continuum (field theory) and the essentially discontinuous interpretation of the elementary structures and processes are fighting for supremacy.

We shall not here discuss questions concerning molecular theory, but shall describe the improvements made in the field theory during this century.

THE THEORY OF RELATIVITY

These all arise from the theory of relativity, which has in the last six months entered its third stage of development. Let us briefly examine the chief points of view belonging to these three stages and their relation to field theory.

[THE SPECIAL THEORY OF RELATIVITY]

The first stage, the special theory of relativity, owes its origin principally to Maxwell's theory of the electromagnetic field. From this, combined with the empirical fact that there does not exist any physically distinguishable state of motion which may be called "absolute rest", arose a new theory of space and time. It is well known that the theory discarded the absolute character of the conception of the simultaneity of two spatially separated events. Well known is also the courage of despair with which some philosophers still defend themselves in a profusion of proud but empty words against this simple theory.

On the other hand, the services tendered by the special theory of relativity to its parent, Maxwell theory of the electromagnetic field, are less adequately recognized. Up to that time the electric field and the magnetic field were regarded as existing separately even if a close causal correlation between the two types of field was provided by Maxwell's field equations. But the special theory of relativity showed that this causal correlation corresponds to an essential identity of the two types of field. In fact, the same condition of space, which in one coordinate system appears as a pure magnetic field, appears simultaneously in another coordinate system in relative motion as an electric field, and vice versa. Relationship of this kind displaying an identity between different conceptions, which therefore reduce the number of independent hypotheses and concepts of field theory and heighten its logical selfcontainedness are a characteristic feature of the theory of relativity. For instance, the special theory also indicated the essential identity of the conceptions' inertial mass and energy. This is all generally known and is only mentioned here in order to emphasize the unitary tendency which dominates the whole development of the theory.

[THE GENERAL THEORY OF RELATIVITY]

We now turn to the second stage in the development of the theory of relativity, the so-called general theory of relativity. This theory also starts from a fact of experience which till then had received no satisfactory interpretation; the equality of inertial and gravitational mass, or, in other the words, the fact known since the days of Galileo and Newton that all bodies fall with equal acceleration in the Earth's gravitational field. The theory uses a special theory as its basis and at the same time modifies it: the recognition that there is no state of motion whatever which is physically privileged - that is, that not only velocity but also acceleration are without absolute significance - forms the starting point of the theory. It then compels a much more profound modification of the conceptions of space and time than were involved in the special theory. For even if the special theory forced us to fuse space and time together to an invisible four-dimensional continuum, yet the Euclidean character of the continuum remained essentially intact in this theory. In the general theory of relativity, this hypothesis regarding the Euclidean character of our space-time continuum had to be abandoned and the latter given the structure of a so-called Riemannian space. Before we attempt to understand what these terms mean, let us recall what this theory accomplished.

It furnished an exact field theory of gravitation and brought the latter into a fully determinate relationship to the metrical properties of the continuum. The theory of gravitation, which until then had not advanced beyond Newton, was thus brought within Faraday's conception of the field in a necessary manner; that is, without any essential arbitrariness in the selection of the field laws. At the same time gravitation and inertia were fused into an essential identity. The confirmation which this theory has received in recent years through the measurement of the deflection of light rays in a gravitational field and the spectroscopic examination of binary stars is well known.

[THE UNITARY FIELD THEORY]

The characteristics which especially distinguish the general theory of relativity and even more the new third stage of the theory, the unitary field theory, from other physical theories are the degree of formal speculation, the slender empirical basis, the boldness in theoretical construction and, finally, the fundamental reliance on the uniformity of the secrets of natural law and their accessibility to the speculative intellect. It is this feature which appears as a weakness to physicists who incline toward realism or positivism, but is especially attractive, nay, fascinating, to the speculative mathematical mind. Meyerson in his brilliant studies on the theory of knowledge justly draws a comparison of the intellectual attitude of the relativity theoretician with that of Descartes, or even of Hegel, without thereby implying the censure which a physicist would read into this. However that may be in the end experience is the only competent judge.

Yet in the meantime one thing may be said in defence of the theory. Advance in scientific knowledge must bring about the result that an increase in formal simplicity can only be won at the cost of an increased distance or gap between the fundamental hypothesis of the theory on the one hand and the directly observed facts on the other hand. Theory is compelled to pass more and more from the inductive to the deductive method, even though the most important demand to be made of every scientific theory will always remain: that it must fit the facts.

We now reach the difficult task of giving to the reader an idea of the methods used in the mathematical construction which led to the general theory of relativity and to the new unitary field theory.

THE PROBLEM STATED

The general problem is: which are the simplest formal structures that can be attributed to a four-dimensional continuum and which are the simplest laws that may be conceived to govern these structures? We then look for the mathematical expression of the physical fields in these formal structures and for the field laws of physics - already known to a certain approximation from earlier researches - in the simplest laws governing this structure.

The conceptions which are used in this connection can be explained just as well in a two-dimensional continuum (a surface) as in the four-dimensional continuum of space and time. Imagine a piece of paper ruled in millimetre squares. What does it mean if I say that the printed surface is two-dimensional? If any point P is marked on the paper, one can define its position by using two numbers. Thus, starting from the bottom left-hand corner, move a pointer toward the right until the lower end of the vertical through the point P is reached. Suppose that in doing this one has passed the lower ends of X vertical (millimetre) lines. Then move the pointer up to the point P passing Y horizontal lines. The point P is then described without ambiguity by the numbers X Y (coordinates). If one had used, instead of ruled millimetre paper, a piece which had been stretched or deformed the same determination could still be carried out: but in this case the lines passed would no longer be horizontals or verticals or even straight lines. The same point would then, of course, yield different numbers, but the possibility of determining a point by means of two numbers (Gaussian coordinates) still remains. Moreover, if P and Q are two points which lie very close to one another, then their coordinates differ only very slightly. When a point can be described by two numbers in this way, we speak of a two-dimensional continuum (surface).

RIEMANNIAN METRIC

Now consider two neighbouring points P, Q, on the surface and a little way off another pair of points P' and Q'. What does it mean to say that the distance P Q is equal to the distance P' Q' ? This statement only has a clear meaning when we have a small measuring rod which we can take from one pair of points to the other and if the result of the comparison is independent of the particular measuring rod selected. If this is so, the magnitudes of the tracts P Q, P' Q' can be compared. If a continuum is of this kind we say it has a metric. Of course, the distance of the two points P Q must depend on the coordinate differences (dx. dy). But the form if this dependence is not known as a priori. It is of the form:

$$ds^2 = g_{11}dx^2 + 2\, g_{11}\, g_{22}\, dx\, dy + g_{22}\, dy^2$$

Then it is called a Riemannian metric. If it is possible to choose the coordinates so that this expression takes the form: $ds^2 = dx^2 + dy^2$ (Pythagoras's theorem), then the continuum is Euclidean (a plane).

Thus it is clear that the Euclidean continuum is a special case of the Riemannian. Inversely, the Riemannian continuum is a metric continuum which is Euclidean in infinitely small regions, but not in finite regions. The quantities g_{11}, g_{12}, and g_{22} describe the metrical properties of the surface; that is, the metrical field.

By making use of empirically known properties of space, especially the law of the propagation of light; it is possible to show that the space-time continuum has a Riemannian metric. The quantities g_{11}, g_{12}, and g_{22}, appertaining to it, determine not only the metric of the continuum, but also the gravitational field. The law governing the gravitational field is found in answer to the question: Which are the simplest mathematical laws to which the metric (that is the g_{11}, g_{12}, and g_{22}) can be subjected? The answer was given by the discovery of the field laws of gravitation, which have proved themselves more accurate than the Newtonian law. This rough outline is intended only to give a general idea of the sense in which I have spoken of the "speculative" methods of the general theory of relativity.

EXPANDING THE THEORY

This theory having brought together the metric and gravitation would have been completely satisfactory if the world had only gravitational fields and no electromagnetic fields. Not it is true that the latter can be included within the general theory of relativity by taking over and appropriately modifying Maxwell's equations of the electromagnetic field, but they do not then appear like the gravitational fields as structural properties of the space-time continuum, but as logically independent constructions. The two types of field are causally linked in this theory, but still not fused to an identity. It can, however, scarcely be imagined that empty space has conditions or states of two essentially different kinds, and it is natural to suspect that this only appears to be so because the structure of the physical continuum is not completely described by the Riemannian metric.

The new unitary field theory removes this fault by displaying both types of field as manifestations of one comprehensive type of spatial structure in the space-time continuum. The stimulus to the new theory arose from the discovery that there exists a structure between the Riemannian space structure and the Euclidean, which is richer in formal relationships than the former but poorer than the latter. Consider a two-dimension Riemannian space in the form of the surface of a hen's egg. Since this surface is embedded in our (accurately enough) Euclidean space, it possesses a Riemannian metric. In fact, it has a perfectly definite meaning to speak of the distance of two neighbouring points P, Q on the surface. Similarly it has, of course, a meaning to say of two such pairs of points (PQ) (P'Q'), at separate parts of the surface of the egg, that the distance PQ is equal to the distance P'Q'. On the other hand, it is impossible now to compare the direction PQ with the direction P'Q'. In particular it is meaningless to demand that P'Q' shall be chosen parallel to PQ. In the corresponding Euclidean geometry of two dimensions, the Euclidean geometry of the plan, directions can be compared and the relationship of parallelism can exist between lines in regions of the plane at any distance from one another (distant parallelism). To this extend the Euclidean continuum is richer in relationships than the Riemannian.

A MATHEMATICAL DISCOVERY

The new unitary field theory is based on the following mathematical discovery: there are continua with a Riemannian metric and distant parallelism which nevertheless are not Euclidean. It is easy to show, for instance, in the case of three-dimensional space, how such a continuum differs from a Euclidean.

First of all, in such a continuum there are lines whose elements are parallel to one another. We shall call those "straight lines". It also has a definite meaning to speak of two parallel straight lines as in the Euclidean case. Now choose two such parallels E_1L_1 and E_2L_2 and mark on each a point P_1, P_2.

On E_1L_1 choose in addition a point Q_1. If we now draw through Q_1 a straight line Q_1-R parallel to the straight line P_1, P_2, then in Euclidean geometry this will cut the straight line E_2L_2; in the geometry now used the line Q_1-R and the line do not in general cut one another. To this extent the geometry now used is not only a specialization of the Riemannian but also a generalization of the Euclidean geometry. My opinion is that our space-time continuum has a structure of the kind here outlined.

The mathematical problem whose solution, in my view, leads to the correct field laws is to be formulated thus: which are the simplest and most natural conditions to which a continuum of this kind can be subjected? The answer to this question which I have attempted to give in a new paper yields unitary field laws for gravitation and electromagnetism.

Glossary of Terms

Some of the terms employed by Dr. Einstein in his article on this page will be recognized by those who recall their elementary physics: others will not be understood except by students of higher mathematics. For the convenience of all readers, the following glossary has been compiled of important terms, names and theories mentioned by Dr. Einstein.

Continuum - A line, straight or curved, is a one-dimension continuum. A surface, flat or curved, is a two-dimension continuum. Space is a three-dimension continuum. Einstein's space time is a four-dimension continuum.
Coulomb Forces - Coulomb, a French physicist (1736-1806), found that two electrified particles attract or repel each other with a force which is directly proportional to the product of their charges and inversely proportional to the square of the distance between them. Such forces are called Coulomb forces.
Deductive Method - Establishing particular facts from general principles or truths.
Diffraction - A deviation of the rays of light from a straight line when they are partially cut off by an obstacle or when they pass near the edges of an opening.
Electromagnetic Field - A portion of space in which electric and magnetic forces exist.
Ether - In physics, ether is a supposed medium which fills all space and through which radiant energy of all kinds - including radio waves, light waves, X rays, cosmic rays is propagated.
Euclid - A Greek mathematician (about 350-300 B.C.) called the "father of geometry".
Faraday - English physicist (1791-1861) and discoverer of electromagnetic induction.
Fourth Dimension - Fourth dimension of space is an assumed dimension whose relation to the recognized dimensions - length, breadth and thickness is analogous to that borne by any one of them to the other two.
Galileo - An Italian physicist and astronomer (1564-1642) inventor of the telescope and discoverer of the moons of Jupiter and the laws of falling bodies.
Gaussian Coordinates - Gauss, a German mathematician (1777-1855). In studying the properties of curved surfaces he used coordinates to latitude and longitude on the surface of a sphere. Such coordinates are called Gaussian coordinates.
Gravitational Field - a portion of space across which heavy bodies attract each other.
Hegel - A German philosopher (1770-1831)
Hertz - A German physicist (1857-1894) who discovered the propagation of electromagnetic waves.
Huygens - Young - Fresnel Wave Theory of Light - Huygens, a Dutch mathematician (1629-1695); Young, an English physicist (1773-1829), and Fresnel, a French physicist (1788-1827), founded the theory that light is propagated by waves.
Inductive Method - The scientific method which attempts to obtain general laws from particular cases.
Inertia - In physics, inertia is that property of matter by virtue of which it persists in its state of rest or of uniform motion, in a straight line, unless some force changes that state.
Interference - In physics, interference is the term used to describe the effect of waves in neutralizing or in reinforcing each other.
Lorentz - A Dutch physicist who developed the theory of electrons.
Maxwell's Field Equations - Maxwell, a Scottish physicist (1831-1879), laid down the electromagnetic theory of light. This theory predicted the effects afterward observed by Hertz. The equations of Maxwell's theory are called the Maxwell field equations.
Metric - A term used to describe a mathematical system of measurement.
Newton's Law - Newton's law of universal gravitation asserts that every particle of matter attracts every other particle of matter with a force which is directly proportional to their masses and inversely proportional to the square of the distance between them.
Ponderable Matter - Matter that has weight.
Quantum Theory - In Atomic Physics, the Quantum theory was founded in 1900 by Planck, A German mathematical Physicist. This theory may be briefly characterized by saying that it considers Atomic phenomena as essentially discontinuous phenomena.
Riemmanian Space and Riemannian Metric - If the properties of two-dimensional space can be described by the formula given in the article, viz:

$$ds^2 = g_{11}dx^2 + 2\ g_{11}\ g_{22}\ dx\ dy + g_{22}\ dy^2$$

The space is said to be a Riemannian space and to have a Riemannian metric. If the two-dimensional space, however, can be described by the simple formula $ds^2 = dx^2 + dy^2$, it is said to be a Euclidean space and to possess Euclidean metric. Riemann was a German mathematician (1836-1876).