

# THE FOUNDATION OF THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

## A. FUNDAMENTAL CONSIDERATIONS ON THE POSTULATE OF RELATIVITY

### § 1. Observations on the Special Theory of Relativity

**T**HE special theory of relativity is based on the following postulate, which is also satisfied by the mechanics of Galileo and Newton.

If a system of co-ordinates  $K$  is chosen so that, in relation to it, physical laws hold good in their simplest form, the *same* laws also hold good in relation to any other system of co-ordinates  $K'$  moving in uniform translation relatively to  $K$ . This postulate we call the "special principle of relativity." The word "special" is meant to intimate that the principle is restricted to the case when  $K'$  has a motion of uniform translation relatively to  $K$ , but that the equivalence of  $K'$  and  $K$  does not extend to the case of non-uniform motion of  $K'$  relatively to  $K$ .

Thus the special theory of relativity does not depart from classical mechanics through the postulate of relativity, but through the postulate of the constancy of the velocity of light *in vacuo*, from which, in combination with the special principle of relativity, there follow, in the well-known way, the relativity of simultaneity, the Lorentzian transformation, and the related laws for the behaviour of moving bodies and clocks.

The modification to which the special theory of relativity has subjected the theory of space and time is indeed far-reaching, but one important point has remained unaffected.

For the laws of geometry, even according to the special theory of relativity, are to be interpreted directly as laws relating to the possible relative positions of solid bodies at rest; and, in a more general way, the laws of kinematics are to be interpreted as laws which describe the relations of measuring bodies and clocks. To two selected material points of a stationary rigid body there always corresponds a distance of quite definite length, which is independent of the locality and orientation of the body, and is also independent of the time. To two selected positions of the hands of a clock at rest relatively to the privileged system of reference there always corresponds an interval of time of a definite length, which is independent of place and time. We shall soon see that the general theory of relativity cannot adhere to this simple physical interpretation of space and time.

## § 2. The Need for an Extension of the Postulate of Relativity

In classical mechanics, and no less in the special theory of relativity, there is an inherent epistemological defect which was, perhaps for the first time, clearly pointed out by Ernst Mach. We will elucidate it by the following example:—Two fluid bodies of the same size and nature hover freely in space at so great a distance from each other and from all other masses that only those gravitational forces need be taken into account which arise from the interaction of different parts of the same body. Let the distance between the two bodies be invariable, and in neither of the bodies let there be any relative movements of the parts with respect to one another. But let either mass, as judged by an observer at rest relatively to the other mass, rotate with constant angular velocity about the line joining the masses. This is a verifiable relative motion of the two bodies. Now let us imagine that each of the bodies has been surveyed by means of measuring instruments at rest relatively to itself, and let the surface of  $S_1$  prove to be a sphere, and that of  $S_2$  an ellipsoid of revolution. Thereupon we put the question—What is the reason for this difference in the two bodies? No answer can

be admitted as epistemologically satisfactory, unless the reason given is an *observable fact of experience*. The law of causality has not the significance of a statement as to the world of experience, except when *observable facts* ultimately appear as causes and effects.

Newtonian mechanics does not give a satisfactory answer to this question. It pronounces as follows:—The laws of mechanics apply to the space  $R_1$ , in respect to which the body  $S_1$  is at rest, but not to the space  $R_2$ , in respect to which the body  $S_2$  is at rest. But the privileged space  $R_1$  of Galileo, thus introduced, is a merely *factitious* cause, and not a thing that can be observed. It is therefore clear that Newton's mechanics does not really satisfy the requirement of causality in the case under consideration, but only apparently does so, since it makes the factitious cause  $R_1$  responsible for the observable difference in the bodies  $S_1$  and  $S_2$ .

The only satisfactory answer must be that the physical system consisting of  $S_1$  and  $S_2$  reveals within itself no imaginable cause to which the differing behaviour of  $S_1$  and  $S_2$  can be referred. The cause must therefore lie *outside* this system. We have to take it that the general laws of motion, which in particular determine the shapes of  $S_1$  and  $S_2$ , must be such that the mechanical behaviour of  $S_1$  and  $S_2$  is partly conditioned, in quite essential respects, by distant masses which we have not included in the system under consideration. These distant masses and their motions relative to  $S_1$  and  $S_2$  must then be regarded as the seat of the causes (which must be susceptible to observation) of the different behaviour of our two bodies  $S_1$  and  $S_2$ . They take over the rôle of the factitious cause  $R_1$ . Of all imaginable spaces  $R_1, R_2$ , etc., in any kind of motion relatively to one another, there is none which we may look upon as privileged *a priori* without reviving the above-mentioned epistemological objection. *The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity.

In addition to this weighty argument from the theory of

\* Of course an answer may be satisfactory from the point of view of epistemology, and yet be unsound physically, if it is in conflict with other experiences.

knowledge, there is a well-known physical fact which favours an extension of the theory of relativity. Let  $K$  be a Galilean system of reference, i.e. a system relatively to which (at least in the four-dimensional region under consideration) a mass, sufficiently distant from other masses, is moving with uniform motion in a straight line. Let  $K'$  be a second system of reference which is moving relatively to  $K$  in *uniformly accelerated* translation. Then, relatively to  $K'$ , a mass sufficiently distant from other masses would have an accelerated motion such that its acceleration and direction of acceleration are independent of the material composition and physical state of the mass.

Does this permit an observer at rest relatively to  $K'$  to infer that he is on a "really" accelerated system of reference? The answer is in the negative; for the above-mentioned relation of freely movable masses to  $K'$  may be interpreted equally well in the following way. The system of reference  $K'$  is unaccelerated, but the space-time territory in question is under the sway of a gravitational field, which generates the accelerated motion of the bodies relatively to  $K'$ .

This view is made possible for us by the teaching of experience as to the existence of a field of force, namely, the gravitational field, which possesses the remarkable property of imparting the same acceleration to all bodies.\* The mechanical behaviour of bodies relatively to  $K'$  is the same as presents itself to experience in the case of systems which we are wont to regard as "stationary" or as "privileged." Therefore, from the physical standpoint, the assumption readily suggests itself that the systems  $K$  and  $K'$  may both with equal right be looked upon as "stationary," that is to say, they have an equal title as systems of reference for the physical description of phenomena.

It will be seen from these reflexions that in pursuing the general theory of relativity we shall be led to a theory of gravitation, since we are able to "produce" a gravitational field merely by changing the system of co-ordinates. It will also be obvious that the principle of the constancy of the velocity of light *in vacuo* must be modified, since we easily

\* Eötvös has proved experimentally that the gravitational field has this

recognize that the path of a ray of light with respect to  $K'$  must in general be curvilinear, if with respect to  $K$  light is propagated in a straight line with a definite constant velocity.

### § 3. The Space-Time Continuum. Requirement of General Co-Variance for the Equations Expressing General Laws of Nature

In classical mechanics, as well as in the special theory of relativity, the co-ordinates of space and time have a direct physical meaning. To say that a point-event has the  $X_1$  co-ordinate  $x_1$  means that the projection of the point-event on the axis of  $X_1$ , determined by rigid rods and in accordance with the rules of Euclidean geometry, is obtained by measuring off a given rod (the unit of length)  $x_1$  times from the origin of co-ordinates along the axis of  $X_1$ . To say that a point-event has the  $X_4$  co-ordinate  $x_4 = t$ , means that a standard clock, made to measure time in a definite unit period, and which is stationary relatively to the system of co-ordinates and practically coincident in space with the point-event,\* will have measured off  $x_4 = t$  periods at the occurrence of the event.

This view of space and time has always been in the minds of physicists, even if, as a rule, they have been unconscious of it. This is clear from the part which these concepts play in physical measurements; it must also have underlain the reader's reflexions on the preceding paragraph (§ 2) for him to connect any meaning with what he there read. But we shall now show that we must put it aside and replace it by a more general view, in order to be able to carry through the postulate of general relativity, if the special theory of relativity applies to the special case of the absence of a gravitational field.

In a space which is free of gravitational fields we introduce a Galilean system of reference  $K(x, y, z, t)$ , and also a system of co-ordinates  $K'(x', y', z', t')$  in uniform rotation relatively to  $K$ . Let the origins of both systems, as well as their axes

\* We assume the possibility of verifying "simultaneity" for events immediately proximate in space, or—to speak more precisely—for immediate proximity or coincidence in space-time, without giving a definition of this fundamental concept.

of  $Z$ , permanently coincide. We shall show that for a space-time measurement in the system  $K'$  the above definition of the physical meaning of lengths and times cannot be maintained. For reasons of symmetry it is clear that a circle around the origin in the  $X, Y$  plane of  $K$  may at the same time be regarded as a circle in the  $X', Y'$  plane of  $K'$ . We suppose that the circumference and diameter of this circle have been measured with a unit measure infinitely small compared with the radius, and that we have the quotient of the two results. If this experiment were performed with a measuring-rod at rest relatively to the Galilean system  $K$ , the quotient would be  $\pi$ . With a measuring-rod at rest relatively to  $K'$ , the quotient would be greater than  $\pi$ . This is readily understood if we envisage the whole process of measuring from the "stationary" system  $K$ , and take into consideration that the measuring-rod applied to the periphery undergoes a Lorentzian contraction, while the one applied along the radius does not. Hence Euclidean geometry does not apply to  $K'$ . The notion of co-ordinates defined above, which presupposes the validity of Euclidean geometry, therefore breaks down in relation to the system  $K'$ . So, too, we are unable to introduce a time corresponding to physical requirements in  $K'$ , indicated by clocks at rest relatively to  $K'$ . To convince ourselves of this impossibility, let us imagine two clocks of identical constitution placed, one at the origin of co-ordinates, and the other at the circumference of the circle, and both envisaged from the "stationary" system  $K$ . By a familiar result of the special theory of relativity, the clock at the circumference—judged from  $K$ —goes more slowly than the other, because the former is in motion and the latter at rest. An observer at the common origin of co-ordinates, capable of observing the clock at the circumference by means of light, would therefore see it lagging behind the clock beside him. As he will not make up his mind to let the velocity of light along the path in question depend explicitly on the time, he will interpret his observations as showing that the clock at the circumference "really" goes more slowly than the clock at the origin. So he will be obliged to define time in such a way that the rate of a clock depends upon where the clock may be.

We therefore reach this result:—In the general theory of relativity, space and time cannot be defined in such a way that differences of the spatial co-ordinates can be directly measured by the unit measuring-rod, or differences in the time co-ordinate by a standard clock.

The method hitherto employed for laying co-ordinates into the space-time continuum in a definite manner thus breaks down, and there seems to be no other way which would allow us to adapt systems of co-ordinates to the four-dimensional universe so that we might expect from their application a particularly simple formulation of the laws of nature. So there is nothing for it but to regard all imaginable systems of co-ordinates, on principle, as equally suitable for the description of nature. This comes to requiring that:—

*The general laws of nature are to be expressed by equations which hold good for all systems of co-ordinates, that is, are co-variant with respect to any substitutions whatever (generally co-variant).*

It is clear that a physical theory which satisfies this postulate will also be suitable for the general postulate of relativity. For the sum of *all* substitutions in any case includes those which correspond to all relative motions of three-dimensional systems of co-ordinates. That this requirement of general co-variance, which takes away from space and time the last remnant of physical objectivity, is a natural one, will be seen from the following reflexion. All our space-time verifications invariably amount to a determination of space-time coincidences. If, for example, events consisted merely in the motion of material points, then ultimately nothing would be observable but the meetings of two or more of these points. Moreover, the results of our measurements are nothing but verifications of such meetings of the material points of our measuring instruments with other material points, coincidences between the hands of a clock and points on the clock dial, and observed point-events happening at the same place at the same time.

The introduction of a system of reference serves no other purpose than to facilitate the description of the totality of such coincidences. We allot to the universe four space-time variables  $x_1, x_2, x_3, x_4$  in such a way that for every point-event

there is a corresponding system of values of the variables  $x_1 \dots x_4$ . To two coincident point-events there corresponds one system of values of the variables  $x_1 \dots x_4$ , i.e. coincidence is characterized by the identity of the co-ordinates. If, in place of the variables  $x_1 \dots x_4$ , we introduce functions of them,  $x'_1, x'_2, x'_3, x'_4$ , as a new system of co-ordinates, so that the systems of values are made to correspond to one another without ambiguity, the equality of all four co-ordinates in the new system will also serve as an expression for the space-time coincidence of the two point-events. As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of co-ordinates to others, that is to say, we arrive at the requirement of general co-variance.

#### § 4. The Relation of the Four Co-ordinates to Measurement in Space and Time

It is not my purpose in this discussion to represent the general theory of relativity as a system that is as simple and logical as possible, and with the minimum number of axioms; but my main object is to develop this theory in such a way that the reader will feel that the path we have entered upon is psychologically the natural one, and that the underlying assumptions will seem to have the highest possible degree of security. With this aim in view let it now be granted that:—

For infinitely small four-dimensional regions the theory of relativity in the restricted sense is appropriate, if the co-ordinates are suitably chosen.

For this purpose we must choose the acceleration of the infinitely small ("local") system of co-ordinates so that no gravitational field occurs; this is possible for an infinitely small region. Let  $X_1, X_2, X_3$ , be the co-ordinates of space, and  $X_4$  the appertaining co-ordinate of time measured in the appropriate unit.\* If a rigid rod is imagined to be given as the unit measure, the co-ordinates, with a given orientation of the system of co-ordinates, have a direct physical meaning

\* The unit of time is to be chosen so that the velocity of light *in vacuo* as measured in the "local" system of co-ordinates is to be equal to unity.



in the sense of the special theory of relativity. By the special theory of relativity the expression

$$ds^2 = - dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \quad . \quad . \quad (1)$$

then has a value which is independent of the orientation of the local system of co-ordinates, and is ascertainable by measurements of space and time. The magnitude of the linear element pertaining to points of the four-dimensional continuum in infinite proximity, we call  $ds$ . If the  $ds$  belonging to the element  $dX_1 \dots dX_4$  is positive, we follow Minkowski in calling it time-like; if it is negative, we call it space-like.

To the "linear element" in question, or to the two infinitely proximate point-events, there will also correspond definite differentials  $dx_1 \dots dx_4$  of the four-dimensional co-ordinates of any chosen system of reference. If this system, as well as the "local" system, is given for the region under consideration, the  $dX_\nu$  will allow themselves to be represented here by definite linear homogeneous expressions of the  $dx_\sigma$  :—

$$dX_\nu = \sum_{\sigma} a_{\nu\sigma} dx_{\sigma} \quad . \quad . \quad . \quad (2)$$

Inserting these expressions in (1), we obtain

$$ds^2 = \sum_{\sigma\tau} g_{\sigma\tau} dx_{\sigma} dx_{\tau}, \quad . \quad . \quad . \quad (3)$$

where the  $g_{\sigma\tau}$  will be functions of the  $x_{\sigma}$ . These can no longer be dependent on the orientation and the state of motion of the "local" system of co-ordinates, for  $ds^2$  is a quantity ascertainable by rod-clock measurement of point-events infinitely proximate in space-time, and defined independently of any particular choice of co-ordinates. The  $g_{\sigma\tau}$  are to be chosen here so that  $g_{\sigma\tau} = g_{\tau\sigma}$ ; the summation is to extend over all values of  $\sigma$  and  $\tau$ , so that the sum consists of  $4 \times 4$  terms, of which twelve are equal in pairs.

The case of the ordinary theory of relativity arises out of the case here considered, if it is possible, by reason of the particular relations of the  $g_{\sigma\tau}$  in a finite region, to choose the system of reference in the finite region in such a way that the  $g_{\sigma\tau}$  assume the constant values

$$\left. \begin{array}{cccc} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{array} \right\} \dots \dots \dots (4)$$

We shall find hereafter that the choice of such co-ordinates is, in general, not possible for a finite region.

From the considerations of § 2 and § 3 it follows that the quantities  $g_{\tau\sigma}$  are to be regarded from the physical standpoint as the quantities which describe the gravitational field in relation to the chosen system of reference. For, if we now assume the special theory of relativity to apply to a certain four-dimensional region with the co-ordinates properly chosen, then the  $g_{\sigma\tau}$  have the values given in (4). A free material point then moves, relatively to this system, with uniform motion in a straight line. Then if we introduce new space-time co-ordinates  $x_1, x_2, x_3, x_4$ , by means of any substitution we choose, the  $g^{\sigma\tau}$  in this new system will no longer be constants, but functions of space and time. At the same time the motion of the free material point will present itself in the new co-ordinates as a curvilinear non-uniform motion, and the law of this motion will be independent of the nature of the moving particle. We shall therefore interpret this motion as a motion under the influence of a gravitational field. We thus find the occurrence of a gravitational field connected with a space-time variability of the  $g_{\sigma}$ . So, too, in the general case, when we are no longer able by a suitable choice of co-ordinates to apply the special theory of relativity to a finite region, we shall hold fast to the view that the  $g_{\sigma\tau}$  describe the gravitational field.

Thus, according to the general theory of relativity, gravitation occupies an exceptional position with regard to other forces, particularly the electromagnetic forces, since the ten functions representing the gravitational field at the same time define the metrical properties of the space measured.

## B. MATHEMATICAL AIDS TO THE FORMULATION OF GENERALLY COVARIANT EQUATIONS

Having seen in the foregoing that the general postulate of relativity leads to the requirement that the equations of

# THE PRINCIPLE OF RELATIVITY

A COLLECTION OF ORIGINAL MEMOIRS ON THE  
SPECIAL AND GENERAL THEORY OF RELATIVITY

BY

H. A. LORENTZ, A. EINSTEIN  
H. MINKOWSKI AND H. WEYL

WITH NOTES BY

A. SOMMERFELD

TRANSLATED BY

W. PERRETT AND G. B. JEFFERY

WITH SEVEN DIAGRAMMS

DOVER PUBLICATIONS, INC.