Albert Einstein

THE PROBLEM OF SPACE, ETHER, AND THE FIELD IN PHYSICS

Albert Einstein

[1879–   ]

The contributions to modern science made by Albert Einstein cannot be fully evaluated in our time. Another age, more imaginatively adapted to his concepts of time-space, may see more than abstract logic in his theory of relativity —a theory which measures distance between events rather than objects and involves both time and space together. It is a system based upon relations, not things, and is mathematically verifiable. The man who, two hundred years after Sir Isaac Newton's death, changed our understanding of the law of gravity and gave to physics an entirely new perspective first published his work on relativity in 1905, when he announced his so-called restricted theory. In 1915 this work was amplified with his generalized theory. Since then corroboration of his new principles has been found in abundance, and now every aspect of the philosophy of science has been changed as a consequence of his investigations. "The Problem of Space, Ether and the Field in Physics" is from Albert Einstein's book, *The World As I See It.*

THE PROBLEM OF SPACE, ETHER, AND THE FIELD IN PHYSICS

ALBERT EINSTEIN

Scientific thought is a development of pre-scientific thought. As the concept of space was already fundamental in the latter, we must begin with the concept of space in pre-scientific thought. There are two ways of regarding concepts, both of which are necessary to understanding. The first is that of logical analysis. It answers the question, How do concepts and judgments depend on each other? In answering it we are on comparatively safe ground. It is the security by which we are so much impressed in mathematics. But this security is purchased at the price of emptiness of content. Concepts can only acquire content when they are connected, however indirectly, with sensible experience. But no logical investigation can reveal this connection; it can only be experienced. And yet it is this connection that determines the cognitive value of systems of concepts.

Take an example. Suppose an archaeologist belonging to a later culture finds a text-book of Euclidean geometry without diagrams. He will discover how the words "point," "straight-line," "plane" are used in the propositions. He will also see how the latter are deduced from each other. He will even be able to frame new propositions according to the known rules.

But the framing of these propositions will remain an empty word-game for him, as long as "point," "straight-line," "plane," etc., convey nothing to him. Only when they do convey something will geometry possess any real content for him. The same will be true of analytical mechanics, and indeed of any exposition of the logically deductive sciences.

What does this talk of "straight-line," "point," "intersection," etc., conveying something to one, mean? It means that one can point to the parts of sensible experience to which those words refer. This extra-logical problem is the essential problem, which the archaeologist will only be able to solve intuitively, by examining his experience and seeing if he can discover anything which corresponds to those primary terms of the theory and the axioms laid down for them. Only in this sense can the question of the nature of a conceptually presented entity be reasonably raised.

With our pre-scientific concepts we are very much in the position of our archaeologist in regard to the ontological problem. We have, so to speak, forgotten what features in the world of experience caused us to frame those concepts, and we have great difficulty in representing the world of experience to ourselves without the spectacles of the old-established conceptual interpretation. There is the further difficulty that our language is compelled to work with words which are inseparably connected with those primitive concepts. These are the obstacles which confront us when we try to describe the essential nature of the pre-scientific concept of space.

One remark about concepts in general, before we turn to the problem of space: concepts have reference to sensible experience, but they are never, in a logical sense, deducible from them. For this reason I have never been able to understand the quest of the *a priori* in the Kantian sense. In any ontological question, the only possible procedure is to seek out those characteristics in the complex of sense experiences to which the concepts refer.

Now as regards the concept of space: this seems to presuppose the concept of the solid object. The nature of the complexes and sense-impressions which are probably responsible for that concept has often been described. The correspondence between certain visual and tactile impressions, the fact that they can be continuously followed out through time, and that the impressions can be repeated at any movement (taste, sight), are some of those characteristics. Once the concept of the solid object is formed in connection with the experiences just mentioned—which concept by no means presupposes that of space or spatial relation—the desire to get an intellectual grasp of the relations of such solid bodies is bound to give rise to concepts which correspond to their spatial relations. Two solid objects may touch one another or be distant from one another. In the latter case, a third body can be inserted between them without altering them in any way, in the former not. These spatial relations are obviously real in the same sense as the bodies themselves. If two bodies are of equal value for the filling of *one* such interval, they will also prove of equal value for the filling of other intervals. The interval is thus shown to be independent of the selection of any special body to fill it; the same is universally true of spatial relations. It is plain that this independence, which is a principal condition of the usefulness of framing purely geometrical concepts, is not necessary *a priori*. In my opinion, this concept of the interval, detached as it is from the selection of any special body to occupy it, is the starting point of the whole concept of space.

Considered, then, from the point of view of sense experience, the development of the concept of space seems, after these brief indications, to conform to the following schema —solid body; spatial relations of solid bodies; interval; space. Looked at in this way, space appears as something real in the same sense as solid bodies.

It is clear that the concept of space as a real thing already existed in the extra-scientific conceptual world. Euclid's mathematics, however, knew nothing of this concept as such; they confined themselves to the concepts of the object, and the spatial relations between objects. The point, the plane, the straight line, length, are solid objects idealised. All spatial relations are reduced to those of contact (the intersection of straight lines and planes, points lying on straight lines, etc.). Space as a continuum does not figure in the conceptual system at all. This concept was first introduced by Descartes, when he described the point-in-space by its co-ordinates. Here for the first time geometrical figures appear, up to a point, as parts of infinite space, which is conceived as a three-dimensional continuum.

The great superiority of the Cartesian treatment of space is by no means confined to the fact that it applies analysis to the purposes of geometry. The main point seems rather to be this:—The geometry of the Greeks prefers certain figures (the straight line, the plane) in geometrical descriptions; other figures (e.g., the ellipse) are only accessible to it because it constructs or defines them with the help of the point, the straight line and the plane. In the Cartesian treatment on the other hand, all surfaces are, in principle, equally represented, without any arbitrary preference for linear figures in the construction of geometry.

In so far as geometry is conceived as the science of laws governing the mutual relations of practically rigid bodies in space, it is to be regarded as the oldest branch of physics. This science was able, as I have already observed, to get along without the concept of space as such, the ideal corporeal forms—point, straight line, plane, length—being sufficient for its needs. On the other hand, space as a whole, as conceived by Descartes, was absolutely necessary to Newtonian physics. For dynamics cannot manage with the concepts of the mass point and the (temporally variable) distance between mass points alone. In Newton's equations of motion the concept of acceleration plays a fundamental part, which cannot be defined by the temporally variable intervals between points alone. Newton's acceleration is only thinkable or definable in relation to space as a whole. Thus to the geometrical reality of the concept of space a new inertia-determining function of space was added. When Newton described space as absolute, he no doubt meant this real significance of space, which made it necessary for him to attribute to it a quite definite state of motion, which yet did not appear to be fully determined by the phenomena of mechanics. This space was conceived as absolute in another sense also; its inertia-determining effect was conceived as autonomous, i.e., not to be influenced by any physical circumstance whatever; it affected masses, but nothing affected it.

And yet in the minds of physicists space remained until the most recent time simply the passive container of all events, playing no part in physical happenings itself. Thought only began to take a new turn with the wave theory of light and the theory of the electromagnetic field of Faraday and Clerk Maxwell. It became clear that there existed in free space conditions which propagated themselves in waves, as well as localised fields which were able to exert force on electrical masses or magnetic poles brought to the spot. Since it would have seemed utterly absurd to the physicists of the nineteenth century to attribute physical functions or states to space itself, they invented a medium pervading the whole of space, on the model of ponderable matter—the ether, which was supposed to act as a vehicle for electro-magnetic phenomena, and hence for those of light also. The states of this medium, imagined as constituting the electro-magnetic fields, were at first thought of mechanically, on the model of the elastic deformations of rigid bodies. But this mechanical theory of the ether was never quite successful and so the idea of a closer explanation of the nature of the etheric fields was given up. The ether thus became a kind of matter whose only function was to act as a substratum for electrical fields which were by their very nature not further analysable. The picture was, then, as follows:—Space is filled by the ether, in which the material corpuscles or atoms of ponderable matter swim; the atomic structure of the latter had been securely established by the turn of the century.

Since the reciprocal action of bodies was supposed to be accomplished through fields, there had also to be a gravitational field in the ether, whose field-law had, however, assumed no clear form at that time. The ether was only accepted as the seat of all operations of force which make themselves effective across space. Since it had been realised that electrical masses in motion produce a magnetic field, whose energy acted as a model for inertia, inertia also appeared as a field-action localised in the ether.

The mechanical properties of the ether were at first a mystery. Then came H. A. Lorentz's great discovery. All the phenomena of electro-magnetism then known could be explained on the basis of two assumptions: that the ether is firmly fixed in space—that is to say, unable to move at all, and that electricity is firmly lodged in the mobile elementary particles. Today his discovery may be expressed as follows: —Physical space and the ether are only different terms for the same thing; fields are physical conditions of space. For if no particular state of motion belongs to the ether, there does not seem to be any ground for introducing it as an entity of a special sort alongside of space. But the physicists were still far removed from such a way of thinking; space was still, for them, a rigid, homogeneous something, susceptible of no change or conditions. Only the genius of Riemann, solitary and uncomprehended, had already won its way by the middle of last century to a new conception of space, in which space was deprived of its rigidity, and in which its power to take part in physical events was recognised as possible. This intellectual achievement commands our admiration all the more for having preceded Faraday's and Clerk Maxwell's field theory of electricity. Then came the special theory of relativity with its recognition of the physical equivalence of all inertial systems. The inseparableness of time and space emerged in connection with electrodynamics, or the law of the propagation of light. Hitherto it had been silently assumed that the four-dimensional continuum of events could be split up into time and space in an objective manner—i.e., that an absolute significance attached to the "now" in the world of events. With the discovery of the relativity of simultaneity, space and time were merged in a single continuum in the same way as the three-dimensions of space had been before. Physical space was thus increased to a four-dimensional space which also included the dimension of time. The four-dimensional space of the special theory of relativity is just as rigid and absolute as Newton's space.

The theory of relativity is a fine example of the fundamental character of the modern development of theoretical science. The hypotheses with which it starts become steadily more abstract and remote from experience. On the other hand it gets nearer to the grand aim of all science, which is to cover the greatest possible number of empirical facts by logical deduction from the smallest possible number of hypotheses or axioms. Meanwhile the train of thought leading from the axioms to the empirical facts or verifiable consequences gets steadily longer and more subtle. The theoretical scientist is compelled in an increasing degree to be guided by purely mathematical, formal considerations in his search for a theory, because the physical experience of the experimenter cannot lift him into the regions of highest abstraction. The predominantly inductive methods appropriate to the youth of science are giving place to tentative deduction. Such a theoretical structure needs to be very thoroughly elaborated before it can lead to conclusions which can be compared with experience. Here too the observed fact is undoubtedly the supreme arbiter; but it cannot pronounce sentence until the wide chasm separating the axioms from their verifiable consequences has been bridged by much intense, hard thinking. The theorist has to set about this Herculean task in the clear consciousness that his efforts may only be destined to deal the death blow to his theory. The theorist who undertakes such a labour should not be carped at as "fanciful"; on the contrary, he should be encouraged to give free rein to his fancy, for there is no other way to the goal. His is no idle day-dreaming, but a search for the logically simplest possibilities and their consequences. This plea was needed in order to make the hearer or reader more ready to follow the ensuing train of ideas with attention; it is the line of thought which has led from the special to the general theory of relativity and thence to its latest offshoot, the unitary field theory. In this exposition the use of mathematical symbols cannot be avoided.

We start with the special theory of relativity. This theory is still based directly on an empirical law, that of the constant velocity of light. Let P be a point in empty space, P′ one separated from it by a length $d\sigma$ and infinitely near to it. Let a flash of light be emitted from P at a time t and reach P′ at a time $t + dt$. Then

$$d\sigma^2 = c^2 dt^2$$

If dx_1, dx_2, dx_3 are the orthogonal projections of $d\sigma$, and the imaginary time co-ordinate $\sqrt{-1}ct = x_4$ is introduced, then the above-mentioned law of the constancy of the propagation of light takes the form

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = 0$$

Since this formula expresses a real situation, we may attribute a real meaning to the quantity ds, even supposing the neighbouring points of the four-dimensional continuum are selected in such a way that the ds belonging to them does not disappear. This is more or less expressed by saying that the four-dimensional space (with imaginary time-co-ordinates) of the special theory of relativity possesses a Euclidean metric.

The fact that such a metric is called Euclidean is connected with the following. The position of such a metric in a three-dimensional continuum is fully equivalent to the positions of the axioms of Euclidean geometry. The defining equation of the metric is thus nothing but the Pythagorean theorem applied to the differentials and the co-ordinates.

Such alteration of the co-ordinates (by transformation) is permitted in the special theory of relativity, since in the new co-ordinates too the magnitude ds^2 (fundamental invariant) is expressed in the new differentials of the co-ordinates by the sum of the squares. Such transformations are called Lorentz transformations.

The leuristic method of the special theory of relativity is characterised by the following principle:—Only those equations are admissible as an expression of natural laws which do not change their form when the co-ordinates are changed by means of a Lorentz transformation (co-variance of equations in relation to Lorentz transformations).

This method led to the discovery of the necessary connection between impulse and energy, the strength of an electric and a magnetic field, electrostatic and electro-dynamic forces, inert mass and energy; and the number of independent concepts and fundamental equations was thereby reduced.

This method pointed beyond itself. Is it true that the equations which express natural laws are co-variant in relation to Lorentz transformations only and not in relation to other transformations? Well, formulated in that way the question really means nothing, since every system of equations can be expressed in general co-ordinates. We must ask, Are not the laws of nature so constituted that they receive no real simplification through the choice of any one *particular* set of co-ordinates?

We will only mention in passing that our empirical principle of the equality of inert and heavy masses prompts us to answer this question in the affirmative. If we elevate the equivalence of all co-ordinate systems for the formulation of natural laws into a principle, we arrive at the general theory of relativity, provided we stick to the law of the constant velocity of light or to the hypothesis of the objective significance of the Euclidean metric at least for infinitely small portions of four-dimensional space.

This means that for finite regions of space the existence (significant for physics) of a general Riemannian metric is presupposed according to the formula

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu}\, dx^\mu\, dx^\nu,$$

whereby the summation is to be extended to all index combinations from 11 to 44.

The structure of such a space differs absolutely radically in *one* respect from that of a Euclidean space. The coefficients $g_{\mu\nu}$ are for the time being any functions whatever of the co-ordinates x_1 to x_4, and the structure of the space is not really determined until these functions $g_{\mu\nu}$ are really known. It is only determined more closely by specifying laws which the metrical field of the $g_{\mu\nu}$ satisfy. On physical grounds this gave rise to the conviction that the metrical field was at the same time the gravitational field.

Since the gravitational field is determined by the configuration of masses and changes with it, the geometric structure of this space is also dependent on physical factors. Thus according to this theory space is—exactly as Riemann guessed—no longer absolute; its structure depends on physical influences. Physical geometry is no longer an isolated self-contained science like the geometry of Euclid.

The problem of gravitation was thus reduced to a mathematical problem: it was required to find the simplest fundamental equations which are co-variant in relation to any transformation of co-ordinates whatever.

I will not speak here of the way this theory has been confirmed by experience, but explain at once why Theory could not rest permanently satisfied with this success. Gravitation had indeed been traced to the structure of space, but besides the gravitational field there is also the electro-magnetic field. This had, to begin with, to be introduced into the theory as an entity independent of gravitation. Additional terms which took account of the existence of the electro-magnetic field had to be included in the fundamental equations for the field. But the idea that there were two structures of space independent of each other, the metric-gravitational and the electro-magnetic, was intolerable to the theoretical spirit. We are forced to the belief that both sorts of field must correspond to verified structure of space.

The "unitary field-theory," which represents itself as a mathematically independent extension of the general theory of relativity, attempts to fulfil this last postulate of the field theory. The formal problem should be put as follows:—Is there a theory of the continuum in which a new structural element appears side by side with the metric such that it forms a single whole together with the metric? If so, what are the simplest field laws to which such a continuum can be made subject? And finally, are these field-laws well fitted to represent the properties of the gravitational field and the electro-magnetic field? Then there is the further question whether the corpuscles (electrons and protons) can be regarded as positions of particularly dense fields, whose movements are determined by the field equations. At present there is only one way of answering the first three questions. The space structure on which it is based may be described as follows, and the description applies equally to a space of any number of dimensions.

Space has a Riemannian metric. This means that the Euclidean geometry holds good in the infinitesimal neighbourhood of every point P. Thus for the neighbourhood of every point P there is a local Cartesian system of co-ordinates, in reference to which the metric is calculated according to the Pythagorean theorem. If we now imagine the length 1 cut off from the positive axes of these local systems, we get the orthogonal "local n-leg." Such a local n-leg is to be found in every other point P′ of space also. Thus, if a linear element (PG or P′G′) starting from the points P or P′, is given, then the magnitude of this linear element can be calculated by the aid of the relevant local n-leg from its local co-ordinates by means of Pythagoras's theorem. There is therefore a definite meaning in speaking of the numerical equality of the linear elements PG and P′G′.

It is essential to observe now that the local orthogonal n-legs are not completely determined by the metric. For we can still select the orientation of the n-legs perfectly freely without causing any alteration in the result of calculating the size of the linear elements according to Pythagoras's theorem. A corollary of this is that in a space whose structure consists exclusively of a Riemannian metric, two linear elements PG and P′G′, can be compared with regard to their magnitude but not their direction; in particular, there is no sort of point in saying that the two linear elements are parallel to one another. In this respect, therefore, the purely metrical (Riemannian) space is less rich in structure than the Euclidean.

Since we are looking for a space which exceeds Riemannian space in wealth of structure, the obvious thing is to enrich Riemannian space by adding the relation of direction or parallelism. Therefore for every direction through P let there be a definite direction through P′, and let this mutual relation be a determinate one. We call the directions thus related to each other "parallel." Let this parallel relation further fulfil the condition of angular uniformity: If PG and PK are two directions in P, P′G′ and P′K′ the corresponding parallel directions through P′, then the angles KPG and K′P′G′ (measurable on Euclidean lines in the local system) should be equal.

The basic space-structure is thereby completely defined. It is most easily described mathematically as follows:—In the definite point P we suppose an orthogonal n-leg with definite, freely chosen orientation. In every other point P′ of space we so orient its local n-leg that its axes are parallel to the corresponding axes at the point P. Given the above structure of space and free choice in the orientation of the n-leg at one point P, all n-legs are thereby completely defined. In the space P let us now imagine any Gaussian system of co-ordinates and that in every point the axes of the n-leg there are projected on to it. This system of n^2 components completely describes the structure of space.

This spatial structure stands, in a sense, midway between the Riemannian and the Euclidean. In contrast to the former, it has room for the straight-line, that is to say a line all of whose elements are parallel to each other in pairs. The geometry here described differs from the Euclidean in the non-existence of the parallelogram. If at the ends P and G of a length PG two equal and parallel lengths PP′ and GG′ are marked off, P′G′ is in general neither equal nor parallel to PG.

The mathematical problem now solved so far is this:—What are the simplest conditions to which a space-structure of the kind described can be subjected? The chief question which still remains to be investigated is this:—To what extent can physical fields and primary entities be represented by solutions, free from singularities, of the equations which answer the former question?