Concerning the Aether

by

Albert Einstein

Einstein, Albert (1924) ‘Über den Äther’,

Verhandlungen der Schweizerischen Naturforschenden Gesellschaft 105:2, 85-93.

Digitised German text: Wikilivres
Scans of German text as published: SEALS (Swiss Electronic Academic Library Service): Retrodigitized Journals

*

When we speak here of aether, we are, of course, not referring to the corporeal aether of mechanical wave-theory that underlies Newtonian mechanics, whose individual points each have a velocity assigned to them. This theoretical construct has, in my opinion, been superseded by the special theory of relativity. **Rather the discussion concerns, much more generally, those things thought of as physically real which, besides ponderable matter consisting of electrical elementary particles, play a role in the causal nexus of physics. Instead of ‘aether’, one could equally well speak of ‘the physical qualities of space’**. Now, it might be claimed that this concept covers all objects of physics, for according to consistent field theory, even ponderable matter, or its constituent elementary particles, are to be understood as fields of some kind or particular ‘states of space’. But it must be admitted that such a view would be premature, since, thus far, all efforts directed toward this goal have foundered. So we are effectively forced by the current state of things to distinguish between matter and aether, even though we may hope that future generations will transcend this dualistic conception and replace it with a unified theory, as the field theoreticians of our day have tried in vain to accomplish.

It is usually believed that aether is foreign to Newtonian physics and that it was only the wave theory of light which introduced the notion of an omnipresent medium influencing, and affected by, physical phenomena. But this is not the case. Newtonian mechanics had its ‘aether’ in the sense indicated, albeit under the name ‘absolute space’. To get a clear understanding of this and, at the same time, to explore more fully the concept of aether, we must take a step back.

We will consider first a branch of physics which makes do without any notion of aether, namely the geometry of Euclid, understood as the study of the possible ways of bringing essentially rigid bodies into contact with each other. (For now, we will set to one side light rays, which may also contribute to the development of geometrical concepts and theorems.) The laws concerning the placement of rigid bodies, excluding relative motion, temperature and the influence of deformations, as laid down in an idealised way in Euclid’s geometry, derive from the concept of a rigid body. Any environmental influence which could be thought of as existing independently of those bodies and as acting on them and influencing the laws governing their placement is unknown to Euclidean geometry. The same holds for the non-Euclidean geometries of constant curvature if these are understood as conceivable laws of nature. It would be different if we were to find ourselves forced to adopt a geometry of variable curvature. This would mean that the laws governing the ways essentially rigid bodies can be brought into contact would be different in different cases, depending on environmental influences. **Here we would have to say that, in the sense we are considering, such a theory would require an aether hypothesis. Its aether would be something every bit as physically real as matter. If the laws of placement were impervious to the influence of physical factors, such as the accumulation and state of motion of bodies in the environment, but irrevocably given, then we would call this aether 'absolute', i.e. by its nature independent of any influence.**

The kinematics, or phoronomy, of classical physics had as little need of an aether as (physically interpreted) Euclidean geometry has. For its laws have a clear physical meaning only if we assume that the special-relativistic influences of motion on rulers and clocks do not exist. Not so in the dynamics of Galileo and Newton. The law of motion 'force equals mass times acceleration', does not consist only of a statement about material systems, not even if, according to Newton's fundamental law of astronomy, the force is expressed at a distance, i.e. by quantities whose 'real definition' [*definitio realis*, a definition in terms of the object's distinguishing properties] can be based on measurements involving rigid bodies. For the 'real definition' of acceleration cannot be completely reduced to observations of rigid bodies and clocks. It cannot be reduced to the measurable distances between the points that make up the mechanical system. Its definition requires also a coordinate system or reference body having some suitable state of motion. If a different coordinate system is chosen, the Newtonian equations do not hold with respect to this new coordinate system. With those equations, the milieu in which the bodies move appears as an implicit, real factor in the laws of motion, alongside the real bodies themselves and the distances that massive bodies define. In contrast to geometry and kinematics, the 'space' of Newton's theory of motion possesses physical reality. **We will call this physical reality which enters the Newtonian law of motion alongside the observable, ponderable real bodies, the aether of mechanics. The occurrence of centrifugal effects with a (rotating) body, whose material points do not change their distances from one another, shows that this aether is not to be understood as a mere hallucination of the Newtonian theory, but rather that i*t corresponds to something real that exists in nature.***

We see that, for Newton, 'space' was something physically real, in spite of the curiously indirect way this real thing reaches our awareness. Ernst Mach, the first after Newton to subject the foundations of mechanics to a deep analysis, perceived this clearly. He sought to escape this hypothesis of the 'mechanical aether' by reducing inertia to immediate interaction between the perceived mass and all other masses of the universe. This view was certainly a logical possibility but, as a theory involving action at a distance, cannot be taken seriously today. The mechanical aether–which Newton called 'absolute space'–must remain for us a physical reality. Of course, one must not be tempted by the expression aether into thinking that, like the physicists of the 19th century, we have in mind something analogous to ponderable matter.

When Newton referred to the space of physics as 'absolute', he was thinking of yet another property of what we call here aether. Every physical thing influences others and is, it its turn, generally influenced by other things. This does not however apply to the aether of Newtonian mechanics. For the inertia-giving property of this aether is, according to classical mechanics, not susceptible to any influence, neither from the configuration of matter nor anything else. Hence the term ‘absolute’.

Only in recent years has it become clear to physicists that the preferred nature of initial systems, as opposed to non-inertial systems, requires a real cause. Viewed historically, the aether hypothesis has emerged in its present form by a process of sublimation from the mechanical aether hypothesis of optics. After long and fruitless efforts, physicists became convinced that light was not to be understood as the motion of an inertial, elastic medium, that the electromagnetic fields of Maxwell’s theory could not be construed as mechanical. So under the pressure of this failure, the electromagnetic fields had gradually come to be regarded as the final, irreducible physical reality, as states of the aether, impervious to further explanation. What remained of the mechanical theory was its definite state of motion; it somehow embodied a state of absolute rest. While at least in Newtonian mechanics all inertial systems were equivalent, it seemed that, in the Maxwell-Lorentz theory, the state of motion of the preferred coordinate system (at rest with respect to the aether) was completely determined. It was accepted implicitly that this preferred coordinate system was also an inertial system, i.e. that the principle of inertia [Newton’s first law] applied relative to the electromagnetic aether.

There was another way too in which the Maxwell-Lorentz theory set back physicists’ basic understanding. Since electromagnetic fields were seen as fundamental, irreducible entities, they seemed destined to rob ponderable masses, possessing inertia, of their primary meaning. It was shown by Maxwell’s equations that a moving, electrically charged body is surrounded by a magnetic field whose energy is, to first approximation, a quadratic function of speed. It seemed only natural to conceive of all kinetic energy as electromagnetic energy. Thus one could hope to reduce mechanics to electromagnetism, since efforts to reduce electromagnetic phenomena to mechanics had failed. Indeed this looked all the more promising as it became apparent that all ponderable matter was composed of electromagnetic elementary particles. But there were two difficulties that could not be overcome. Firstly the Maxwell-Loretz equations could not explain how the electric charge constituting an electrical elementary particle can exist in equilibrium in spite of the forces of electrostatic repulsion. Secondly electromagnetic theory could not give a reasonably natural and satisfactory explanation of gravitation. Nevertheless the results that electromagnetic theory achieved for physics were so significant they came to be regarded as a completely secured possession, indeed as its most firmly established success.

The Maxwell-Lorentz theory eventually influenced our view of the theoretical basis to the extent that it led to the creation of the special theory of relativity. It was recognised that the equations of electromagnetism did not, in fact, single out one particular state of motion, but rather that, in accordance with these equations, just as with those of classical mechanics, there exists an infinite multitude of coordinate systems in mutually equivalent states of motion, providing the appropriate transformation formulas are used for the spatial and temporal coordinates. It is well known that this realisation entailed a profound modification, not only in our ideas about space and time, but also to kinematics and dynamics. No longer was a special state of motion to be ascribed to the electromagnetic aether. Now, like the aether of classical mechanics, it resulted not in the favoring of a particular state of motion, only the favoring of a particular state of acceleration. Because it was no longer possible to speak, in any absolute sense, of simultaneous states at different locations in the aether, the aether became, as it were, four dimensional, since there was no objective way of ordering its states by time alone. According to special relativity too, the aether was absolute, since its influence on inertia and the propogation of light was thought of as being itself independent of physical influence. While classical physics took it for granted that the geometry of bodies was independent of their state of motion, the special theory of relativity stated that the laws of Euclidean geometry only apply to the positioning of bodies at rest with respect to one another when these bodies are at rest with respect to an inertial coordinate system. [1] This can be easily concluded from the so-called Lorentz contraction. Thus geometry, like dynamics, came to depend on the aether.

The general theory of relativity rectified a mischief of classical dynamics. According to the latter, inertia and gravity appear as quite different, mutually independent phenomena, even though they both depend on the same quantity, mass. The theory of relativity resolved this problem by establishing the behaviour of the electrically neutral point-mass by the law of the geodetic line, according to which inertial and gravitational effects are no longer considered as separate. In doing so, it attached characteristics to the aether which vary from point to point, determining the metric and the dynamic behaviour of material points, and determined, in their turn, by physical factors, namely the distribution of mass/energy.

Thus the aether of general relativity differs from those of classical mechanics and special relativity in that it is not ‘absolute’ but determined, in its locally variable characteristics, by ponderable matter. This determination is a complete one if the universe is finite and closed. That there are, in general relativity, no preferred spacetime coordinates uniquely associated with the metric is more characteristic of its mathematical form than its physical framework.

Even using mathematical apparatus of general relativity it has not been possible to reduce all of the inertia of mass to electromagnetic fields, or to fields in general. Neither are we yet, in my view, at the point of formally incorporating the electromagnetic forces into the scheme of general relativity. O**n the one hand, the metric tensor, which codetermines the phenomena of gravitation and inertia and, on the other, t*****he tensor of the electromagnetic field appear still as different expressions of the state of the aether,*** **whose logical independence one is inclined to attribute rather to the** ***incompleteness of our theoretical ediface*** **than to a complex structure of reality.**

It is true that Weyl and Eddington have, by a generalisation of Riemannian geometry, found a mathematical system, in which both kinds of field appear to be unified under a single perspective. But the simplest field laws which that theory provides seem to me not to advance physical insight. On the whole, we seem to be much further now from an understanding of the fundamental laws of electromagnetism than we did at the beginning of this century. As justification for this opinion, I should here like to briefly refer to the problem of the magnetic fields of the earth and the sun, and also to the problem of light quanta, which problems have some bearing on the gross and fine structure of the electromagnetic field.

The earth and sun possess magnetic fields whose orientation and sense are closely related to the spin axes of these bodies. According to Maxwell’s theory, these fields may be due to electric currents which flow in the opposite direction to the rotation of the earth and sun about their axes. Even sunspots, which there are good grounds to think of as vortices, posses analogous, and very powerful, magnetic fields. But it is hardly conceivable that, in all these cases, circuits or convection currents of sufficient strength are actually present. Rather it looks as if cyclic motion of neutral masses generated magnetic fields. Neither Maxwell's theory as originally conceived nor as extended in general relativity predict field generation of that sort. Here nature seems to point us toward some fundamental connection, not yet understood.[2]

If the case we have just discussed is one that field theory, in its current form, seems not yet able to address, the facts and ideas subsumed under quantum theory threaten to the blow the edifice of field theory to bits. Specifically, we find increasing arguments suggesting that the quanta of light are to be understood as physical reality, and that the electromagnetic field cannot be seen as the final reality to which all other physical objects can be reduced. As Planck's formula had already shown that the transmission of energy and momentum by radiation happens as if the latter consisted of particles moving at the speed of light, , with energy so Compton demonstrated, by his research into the scattering of X-rays by matter that scattering events occur in which quanta of light collide with electrons and transmit to them a portion of their energy, as a result of which the quanta of light undergo a change of energy and direction. It is at least a fact that X-rays experience such changes in frequency on scattering (in agreement with the predictions of Debye and Compton) as quantum theory demands.

Recently there has appeared work by the Indian physicist Bose on the derivation of Planck's formula which is of particular significance to our theoretical understanding for the following reasons: hitherto all complete derivations of Planck's formula made some use of the hypothesis of the wave structure of radiation. So, for example, in the well-known Ehrenfest-Debye derivation, the factor in this formula was deduced by counting the eigenvibrations of the cavity belonging to the frequency range . Bose replaces this derivation based on the ideas of wave theory with a gas-theoretical calculation which he applies to a quantum of light conceived of like some sort of molecule present in the cavity. This raises the question of whether it might perhaps also be possible to link the phenomena of diffraction and interference to quantum theory in such a way that the field-like concepts of the theory are presented only as expressions of the interaction between quanta, so that independent physical reality would no longer be ascribed to the fields.

The important fact that the radiation emitted is not, according to Bohr's frequency theory, determined by electrically charged masses which periodically cycle through occurrences of the same frequency can only strengthen this doubt of ours as to the independent reality of the wave field.

But even if these possibilities do mature into an actual theory, we will not be able to do without the aether in theoretical physics, that is, a continuum endowed with physical properties; for general relativity, to whose fundamental viewpoints physicists will always hold fast, rules out direct action at a distance. ***But every theory of local action assumes continuous fields, and thus also the existence of an 'aether'.***