INVARIANCE IN PHYSICAL THEORY

EUGENE P. WIGNER

Thomas D. Jones Professor of Mathematical Physics, Princeton University

INITIAL CONDITIONS, LAWS OF NATURE, INVARIANCE

THE world is very complicated and it is clearly impossible for the human mind to understand it completely. It has therefore devised an artifice which permits the complicated nature of the world to be blamed on something which is called accidental and thus permits him to abstract a domain in which simple laws can be found. The complications are called initial conditions, the domain of regularities, laws of nature. Unnatural as such a division of the world's structure may appear from a very detached point of view, and probable though it is that the possibility of such a division has its own limits,[1] the underlying abstraction is probably one of the most fruitful ones that human mind has made. It has made science possible.

The possibility of abstracting laws of motion from the chaotic set of events that surround us is based on two circumstances. First, in many cases a set of initial conditions can be isolated which is not too large a set and, in spite of this, contains all the relevant conditions for the events on which one focuses one's attention. In the classic example of the falling body, one can disregard almost everything except the initial position and velocity of the falling body; its behavior will be the same and independent of the degree of illumination, the neighborhood of other objects, their temperature, etc. The isolation of the set of conditions which do influence the experiment is by no means a trivial problem. On the contrary, it is a large fraction of the art of the experimenter and on our occasional trips through laboratories all of us theoreticians have been periodically impressed by the difficulties of this art.

However, the possibility of isolating the relevant initial conditions would not in itself make possible the discovery of laws of nature. It is, rather, also essential that, given the same essential initial conditions, the result will be the same no matter where and when we realize these. This principle can be formulated, in the language of initial conditions, as the statement that the absolute position and the absolute time are never essential initial conditions. The statement, that absolute time and position are never essential initial conditions, is the first and perhaps the most important theorem of invariance in physics. If it were not for it, it might have been impossible for us to discover laws of nature.

The above invariance is called in modern mathematical parlance invariance with respect to displacement in time and space. Again, it may be well to remember that this invariance may have limitations. If the universe should turn out to be grossly inhomogeneous, the laws of nature on the fringes of the universe may be quite different from those which we are studying; and it is not impossible that an experimenter inside a closed room is in principle able to ascertain whether he is in the midst, or near the fringes, of the universe, whether he lives in an early epoch of the expansion of the universe, or at an advanced stage of this process. The postulate of the invariance with respect to displacement in space and time disregards this possibility and its application on the cosmological scale virtually presupposes a homogeneous and stationary universe. Present evidence clearly points to the approximate nature of the latter assumption.

INVARIANCE

What are the other laws of invariance? One can distinguish between two types of laws of invariance: the older ones which found their perfect, and perhaps final, formulation in the special theory of relativity, and the new one, yet incompletely

[1] The artificial nature of the division of information into "initial conditions" and "laws of nature" is perhaps most evident in the realm of cosmology. Equations of motion which purport to be able to predict the future of a universe from an arbitrary present state clearly cannot have an empirical basis. It is, in fact, impossible to adduce reasons against the assumption that the laws of nature would be different even in small domains if the universe had a radically different structure. One cannot help agreeing to a certain degree with E. A. Milne who reminds us (*Kinematic relativity,* Oxford Univ. Press, 1948, page 4) that, according to Mach, the laws of nature are a consequence of the contents of the universe. The remarkable fact is that this point of view could be so successfully disregarded and that the distinction between initial conditions and laws of nature has proved so fruitful.

PROCEEDINGS OF THE AMERICAN PHILOSOPHICAL SOCIETY, VOL. 93, NO. 7, DECEMBER, 1949

understood, which the general theory of relativity brought us.

The older theories of invariance postulate, in addition to the irrelevance of the absolute position and time of an event, the irrelevance of its orientation and finally, the irrelevance of its state of motion, as long as this remains uniform, free of rotation, and on a straight line. The former theorems are geometrical in nature and appear to be so self-evident that they were not formulated clearly and directly until about the turn of the last century. The last one, the irrelevance of the state of motion is far from self-evident, as all of us know who have tried to explain it to a layman. There would be no such principle of invariance if Newton's second law of motion read "All bodies persist in their state of rest unless acted upon by an external force"; on the contrary, the scope of this invariance could be extended considerably if the bodies maintained their state of acceleration rather than their velocity in the absence of an external force. It is fitting that this principle was first enunciated, in full clarity, by Newton in his *Principia.*

The fact that the older principles of invariance are the products of experience rather than *a priori* truths can also be illustrated by our gradual abandonment of a very plausible principle, the principle of similitude. This principle, formulated perhaps most clearly by Fourier, demands that physical experiments could be scaled; that the absolute magnitude of objects be irrelevant from the point of view of their behavior on the proper scale. The existence of atoms, of an elementary charge, and of a limiting velocity spelled the doom of this principle.

The formulae describing what I am calling the older principles of invariance were first given completely by Poincaré who derived them from the equations of electrodynamics. He also recognized the group property of the older principles of invariance and named the underlying group after Lorentz. The significance and general validity of these principles were recognized, however, only by Einstein. His papers on special relativity also mark the reversal of a trend: until then, the principles of invariance were derived from the laws of motion. Einstein's work established the older principles of invariance so firmly that we have to be reminded that they are based only on experience. It is now natural for us to try to derive the laws of nature and to test their validity by means of the laws of invariance, rather than to derive the laws of invariance from what we believe to be the laws of nature.

The general theory of relativity is the next milestone in the history of invariance. The fact that it is the first attempt to derive a law of nature by selecting the simplest invariant equation would in itself justify the epithet. More important, in my opinion, is that the general theory of relativity attempts to give the range of the validity of the older theorems of invariance and to replace them with a single, more general theorem. The limitation of the older theorems of invariance is given by the structure of space which manifests itself in a variable curvature. Since the curvature is, in principle, observable, a displacement from a region of low curvature to one with a high curvature does not leave the laws of nature invariant. It is true that the old fashioned physicist can always blame the differences in the laws of nature, as they are valid for different points of the universe, on the absence or proximity of masses. This, however, restores the general validity of the older invariances only by making them meaningless. Clearly, if two points in space time are equivalent only if they are surrounded by the same distribution of masses, their equivalence will be the exception rather than the rule.

The new principle of invariance which the general theory of relativity substitutes for the older ones is that all actions are transmitted by fields which transmit the perturbations from point to point. Expressed more phenomenologically: the events in one part of space depend only on the fields, i.e., on the measurable quantities, in the neighborhood of that part of space; the effect of events outside moves in only with a finite velocity. This postulate of invariance is much bolder, and has much less artificiality than the older postulate of invariance with respect to the inhomogeneous Lorentz group. The above formulation is a little more phenomenological than the customary one. The customary requirement of invariance with respect to all differentiable coordinate transformations is, however, included in it. Both postulates express the fact that the laws of physics and of geometry involve only local measurements such as can be expressed by differential equations. In particular, the definition of a preferred Galilean coordinate system, by reference to other, distant Galilean coordinate systems, is barred by the postulate that all the information which is necessary to describe the immediate future of the region in question can be obtained by local measurements. Hence information relating to distant points cannot add anything relevant to the knowledge of local conditions as would be the case if they would enable one to define preferred coordinate systems.

INVARIANCE IN QUANTUM MECHANICS

When the great paradoxons of atomic physics first became apparent about thirty years ago, it was easy to despair to such a degree of our ability to understand the laws of physics as to propose throwing into the winds all laws of physics, excepting the conservation laws for energy and momenta. It was, in fact, Einstein who recommended such a procedure.[2]

The efforts of the past thirty years culminated in having accomplished just that: we now believe that we have a consistent theory of atomic processes, consistent with the older concepts of space and time, and of invariance. This theory is based on an analysis of the measuring process, carried out principally by Heisenberg and Bohr, which emphasizes the effect of the measurement on the measured object. It is thus contradictory to the simple concept of measuring out the field, the concept which underlies the customary formulation of general relativity. In particular, the measurement of the curvature of space caused by a single particle could hardly be carried out without creating new fields which are many billion times greater than the field under investigation.[3]

Very little effort has been made so far to modify the concepts of the general theory of relativity with an appreciation of the effect of the act of measurement on the object of the measurement. However, the older principles of invariance are in harmony with quantum mechanics and this harmony is more complete, the interdependence of quantum equations and the theory of their invariance is more intimate, than it was in prequantum theory.

Let me first stress the points of similarity between the role of invariance in classical and quantum theories. The principles of invariance have a dual function in both theories. On the one hand, they give a necessary condition which all fundamental equations must satisfy: the irrelevant initial conditions must not enter in a relevant fashion into the results of the theory. Second, once the fundamental equations are given, the principles of invariance furnish, in the form of conservation laws and otherwise, powerful assistance toward their solution. The conservation laws for linear momentum and energy, for angular momentum and the motion of the center of mass, can be derived both in classical theory and in quantum mechanics from the invariance of the equations with respect to infinitesimal displacements and rotations in space-time.[4]

However, with these points of analogy, the similarity between the roles of invariance in classical and in quantum physics is pretty much at an end. The reason is, fundamentally, that the variety of states is much greater in quantum theory than in classical physics and that there is, on the other hand, the principle of superposition to provide a structure for the greatly increased manifold of quantum mechanical states. The principle of superposition renders possible the definition of states the transformation properties of which are particularly simple. It can in fact be shown that every state of any quantum mechanical system, no matter what type of interactions are present, can be considered as a superposition of states of elementary systems. The elementary systems correspond mathematically to irreducible representations of the Lorentz group and as such can be enumerated. Since the equations of motion of the states of elementary systems are completely determined by their invariance properties, every state is a linear combination of states the history of which is completely known. However, in the description by irreducible states, the form of almost all physically important operators remains unknown and, in fact, depends on the system, the type of interactions, etc. This leads to a rather strange dilemma; in the customary description the form of the physically important operators is known but the time dependence of the states is unpredictable or difficult to calculate. In the description just mentioned, the situation is opposite; the time dependence of the states follows from the invariance properties, but the form of the physically important operators is hard to establish. There is one exception to this; the

[2] H. Poincaré, Dynamics of electrons, *Comptes Rendus* **140**: 1504, 1905; Sur la dynamique de l'électron, *Circolo Mat. Palermo Rend.* **21**: 129, 1906.

[3] An interesting problem in this connection was broached recently by M. F. M. Osborne, Quantum theory restrictions on the general theory of relativity, *Bull. Amer. Phys. Soc.* **24** (2) (Berkeley Meeting) Paper A–3, 1949.

[4] In classical theory, this observation is due to F. Klein's school. *Cf.* also F. Engel, Uber die zehn allgemeinen Integrale der klassischen Mechanik, *K. Gesell. der Wiss. zu Göttingen, Nach.*, 1916, p. 270; also G. Hamel, Die Lagrange-Eulerschen Gleichungen der Mechanik, *Ztschr. f. Math. u. Physik* **50**: 1, 1904, and E. Bessel-Hagen, Uber die Erhaltungssätze der Elektrodynamik, *Math. Ann.* **84**: 258, 1921.

states of elementary particles are formed by the superposition of the states of a single invariant set. As a result, the possible equations of elementary particles can easily be enumerated and some progress has been made recently also toward the invariant theoretic determination of the operators for the most important physical quantities. The property which makes a particle elementary in the sense of the above statement is that it shall have no internal coordinate, which would permit an invariant division of its states into two or more groups. It is certainly no accident that all elementary particles, including the light quantum, obey irreducible equations and hence form elementary systems in the above sense. Since the rigid body is what may be considered classical mechanic's closest analogue to an elementary particle, the group theoretical description of the motion of a rigid body must be considered the closest analogue to the above result.

The second point to which I wish to draw attention in the comparison of quantum and pre-quantum theories concerns the significance of transformations of invariance, such as reflections, which cannot be generated by infinitesimal elements. These had very little role in the classical theory but prove their value both in the discussion of fundamental equations, and also in the attempts to solve these. Into the former category belongs for instance the observation that the theory [5] which identifies the neutrino with the antineutrino, by attributing to the inversion of space coordinates a non-linear operation involving transition to the conjugate complex wave function, cannot be welded into a theory which describes also particles of the conventional type.[6] The applications of the reflection invariance for facilitating the solution of the fundamental equations are even more obvious. They lead for instance to the concept of Laporte's parity quantum member—one of the most important concepts of spectroscopy.

Less specifically, but perhaps not less accurately, one can speak of the general impression of quantum mechanics, and the theory of the invariance of its equations, forming an inseparable entity, almost to the degree to which this is true in the general theory of relativity. Schwinger's quantum electrodynamics gives the latest and starkest manifestation of this situation: his theory cannot be formulated at all without developing, unified with it, its theory of invariance. Furthermore, one is inclined to believe that this union is the most important success of the theory; that even the explanation of definite and previously unexpected experimental phenomena is less important to us than the knowledge that we can, in general, carry out our calculations of physical phenomena in an invariant fashion, obtaining the same results if we start with only irrelevantly different initial conditions.

CONSERVATION OF ELECTRICAL CHARGE

My account of the role of invariance in quantum mechanics would remain grossly incomplete if I did not mention a dissonant sound in the harmony of quantum mechanics and the older theorems of invariance. This is the conservation law for the electrical charge. While the conservation laws for all other quantities, such as energy or angular momentum, follow in a natural way from the principles of invariance, the conservation law for electric charge so far has defied all attempts to place it on an equally general basis. The situation was, of course, the same in classical mechanics but the simplicity of the connection between invariance and the ordinary conservation laws makes the situation even more conspicuous in quantum mechanics.

A short description of the derivation of the usual conservation laws will make this perhaps more evident than an abstract discussion. In order to derive the conservation law for linear momentum, one first constructs a state in which one component, say the x component of the linear momentum, has a definite value p. For this purpose, one chooses an arbitrary state $_0$ of the system for which one wishes to show the conservation theorem and constructs all states φ_a obtained by displacing the system in the state φ_0 by a in the x direction. One then considers the superposition of the states φ_a with the coefficients e^{-ipa}

$$\Phi_p = \int_{-\infty}^{\infty} \varphi_a e^{-ipa} da$$

This state has the property that a further displacement by b

$$\int_{-\infty}^{\infty} \varphi_{a+b} e^{-ipa} da = \int_{-\infty}^{\infty} \varphi_c e^{-ip(c-b)} dc = e^{ipb}\Phi_p$$

just multiplies it with e^{ipb}. It is called a pure state with momentum component p in the x direction. The property of Φ_p, of being multiplied by

[5] *Cf.* H. Weyl, Elektron und gravitation I, *Ztschr. f. Physik* **56**: 330, 1929.

[6] Another very interesting set of examples has been given recently by T. Okayama, On the mesic charge, *Phys. Rev.* **75**: 308, 1949.

e^{ipb} upon displacement by b, will not be lost in time: if φ_0 goes over, after some time, into ψ_0, the state φ_a will go over into the ψ_a which results from ψ_0 by displacement by a. This follows from the invariance of the equations of motion with respect to displacements. As a result of this and the linearity of the equations of motion, Φ_p will go over at the time in question into

$$\Psi_p = \int_{-\infty}^{\infty} \psi_a e^{-ipa} da$$

which also is multiplied by e^{ipb} upon displacement by b. This property, which characterizes the state with momentum p, is not lost in the course of time and this constitutes the principle of conservation of linear momentum.

Similar considerations involving the other principles of invariance lead to the other conservation laws. Furthermore, the quantization and the possible values of the quantized quantities also emerge naturally from the above consideration. Thus the quantization of the angular momentum is the result of the condition that rotation by 2π always restores the system to its original state.

No consideration similar in generality and simplicity to the above one is known which would explain the conservation law for electric charges. One can borrow the following argument from classical theory:[7] Suppose we could create charges by some process in a closed system. Let us put then this closed system into a Faraday cage, charge the cage, and create the charge in the closed system. A certain energy E will be necessary for this process. However, inasmuch as no physical phenomenon depends on the absolute value of the potential, the amount of energy E cannot depend on the potential of the Faraday cage inside of which the charge is created. Let us then take our closed system out of the Faraday cage and move it away from it, thereby obtaining a certain amount of work W. Let us then reverse the process which led to the creation of the charge and gain an amount E of energy which is equal to the amount of energy expended in the first place, since the process in a closed system must not depend on the absolute value of the electric potential at which that system is. We now can replace the discharged system into the Faraday cage without the expenditure of any work and have carried out a cycle which resulted in a net gain W of work. This is impossible according to the first law and shows that one of our assumptions must have been faulty. It is the assumption that electric charges can be created in a closed system.

The above argument shows the connection between the conservation law for electric charges and the assumption of the irrelevance of the absolute magnitude of the electric potential. It has been translated into quantum mechanics and has been given a much more elegant and general form.[8] Nevertheless, it falls short of the consideration leading to the other conservation laws in convincing force and, certainly, in its failure to account for the quantization of the electric charge.

The lack of full clarity concerning the foundation of the conservation law for charges raises several important points. Is our present scheme of quantum mechanics incomplete in some fundamental respect? In particular, is the Hilbert space with complex coordinates the proper framework for describing state vectors? Would the use of more general hypercomplex wave functions give essentially different results? But the most important question is, undoubtedly: is the existence of a conservation law a particular feature of the electromagnetic type of interaction or are we going to encounter, or perhaps have we already encountered,[9] similar conservation laws for other types of interaction?

[7] This point was emphasized by J. R. Oppenheimer during the discussion which followed the presentation of this paper.

[8] F. London, Quantenmechanische Deutung der Theorie von Weyl, *Ztschr. f. Physik* **42**: 375, 1927; *Cf.* also H. Weyl, *loc. cit.*.

[9] It is conceivable, for instance, that a conservation law for the number of heavy particles (protons and neutrons) is responsible for the stability of the protons in the same way as the conservation law for charges is responsible for the stability of the electron. Without the conservation law in question, the proton could disintegrate, under emission of a light quantum, into a positron, just as the electron could disintegrate, were it not for the conservation law for the electric charge, into a light quantum and a neutrino. The Gedanken experiment which led to the conservation law for charges would assume the following, admittedly somewhat vague form, if one wanted to prove a conservation law for the number of nucleons: Assuming that there is no such conservation law, two nucleons could first be created at a distance from each other which is large compared with the range of nuclear forces. An amount E of energy would be needed for this. The nucleons then could be permitted to approach each other, furnishing the amount W of work. Finally, they would be permitted to annihilate which would again release the energy E first expended. A net gain W in energy would result. The impossibility to perform the above operations may, of course, be connected with many physical phenomena, such as the im-

possibilty of localizing sufficiently accurately the systems in which the nucleons are to be created (i.e., the existence of a fundamental length). The impossibility may also be the consequence of the dependence of the energy E, which is necessary to create the nucleons, on the absolute value of the nuclear potential. The point of view which we wish to represent is, however, that the impossibility of the creation of nucleons (without creating antinucleons) is the real resolution of the paradoxon. It may be mentioned, as a third point of similarity between the two conservation laws, that there is evidence, although contested by some recent experiments, that the "mesonic charge" of all nucleons is the same. If this should prove to be true, it would be evidence for the quantization of the mesonic charge. This quantization would be analogous to the well known quantization of the electric charge.

Relativity theory, to the celebration of which the present paper is intended to contribute, has enriched physics in two ways. It has resolved acute difficulties, presented by the Michelson-Morley, the Fizeau, the Trouton-Noble and other experiments. It has done this by a profound analysis of the space-time concept and its results in this connection are part of the store of knowledge of all physicists. Even more lasting and more subtle is probably the contribution which relativity theory has made indirectly. Most important among the indirect contributions of the theory of relativity was its demonstration for the need and fruitfulness of the analysis of apparently well established concepts, such as have formed a habit of thought for many generations. Its fostering the emergence of the importance of the concept of invariance, its enlarging the scope of this concept, can, I believe, justly claim second position.

It is a pleasure to acknowledge Dr. V. Bargmann's critical comments and remarks on the present paper.