EUGENE P. WIGNER

Events, laws of nature, and invariance principles

Nobel Lecture, December 12, 1963

It is a great and unexpected honor to have the opportunity to speak here today. Six years ago, Yang and Lee spoke here, reviewing symmetry principles in general and their discovery of the violation of the parity principle in particular[1]. There is little point in repeating what they said, on the history of the invariance principles, or on my own contribution to these which they, naturally, exaggerated. What I would like to discuss instead is the general role of symmetry and invariance principles in physics, both modern and classical. More precisely, I would like to discuss the relation between three categories which play a fundamental role in all natural sciences: events, which are the raw materials for the second category, the laws of nature, and symmetry principles for which I would like to support the thesis that the laws of nature form the raw material.

Events and Laws of Nature

It is often said that the objective of physics is the explanation of nature, or at least of inanimate nature. What do we mean by explanation? It is the establishment of a few simple principles which describe the properties of what is to be explained. If we understand something, its behavior, that is the events which it presents, should not produce any surprises for us. We should always have the impression that it could not be otherwise.

It is clear that, in this sense, physics does not endeavor to explain nature. In fact, the great success of physics is due to a restriction of its objectives: it only endeavors to explain the *regularities* in the behavior of objects. This renunciation of the broader aim, and the specification of the domain for which an explanation can be sought, now appears to us an obvious necessity. In fact, the specification of the explainable may have been the greatest discovery of physics so far. It does not seem easy to find its inventor, or to give the exact date of its origin. Kepler still tried to find exact rules for the magnitude of the planetary orbits, similar to his laws of planetary motion. Newton already

realized that physics would deal, for a long time, only with the explanation of those of the regularities discovered by Kepler which we now call Kepler's laws[2].

The regularities in the phenomena which physical science endeavors to uncover are called the laws of nature. The name is actually very appropriate. Just as legal laws regulate actions and behavior under certain conditions, but do not try to regulate all actions and behavior, the laws of physics also determine the behavior of its objects of interest only under certain well defined conditions, but leave much freedom otherwise. The elements of the behavior which are not specified by the laws of nature are called initial conditions. These, then, together with the laws of nature, specify the behavior as far as it can be specified at all: if a further specification were possible, this specification would be considered as an added initial condition. As is well known, before the advent of quantum theory, it was believed that a complete description of the behavior of an object is possible so that, if classical theory were valid, the initial conditions and the laws of nature together would completely determine the behavior of an object.

The preceding statement is a definition of the term «initial condition». Because of its somewhat unusual nature, it may be worthwhile to illustrate this on an example. Suppose we did not know Newton's equation for the motion of stars and planets

$$\ddot{\mathfrak{r}}_i = G\,\Sigma' M_j\,\frac{\mathfrak{r}_{ij}}{r_{ij}^3} \qquad \mathfrak{r}_{ij} = \mathfrak{r}_j - \mathfrak{r}_i \tag{1}$$

but had found only the equation determining the third derivative of the position

$$\dddot{\mathfrak{r}}_i = G\Sigma' M_j\,\frac{\dot{\mathfrak{r}}_{ij}(\mathfrak{r}_{ij}\cdot\mathfrak{r}_{ij}) - 3\,\mathfrak{r}_{ij}(\dot{\mathfrak{r}}_{ij}\cdot\mathfrak{r}_{ij})}{r_{ij}^5} \tag{2}$$

More generally, if the forces F_i are non-gravitational, one would have written

$$M_i\,\ddot{\mathfrak{r}}_i = \dot{\mathfrak{r}}_i\cdot \mathrm{grad} F_i + \dot{F}_i \tag{2a}$$

The initial conditions then would contain not only all the $\mathfrak{r}_i$ and $\dot{\mathfrak{r}}_i$, but also the $\ddot{\mathfrak{r}}_i$. These data, together with the «equation of motion» (2), would then determine the future behavior of the system just as $r_i, \dot{r}_i$ and (1) determines it. The fact that initial conditions and laws of nature completely determine the behavior is, similarly, true in any causal theory.

The surprising discovery of Newton's age is just the clear separation of laws

of nature on the one hand and initial conditions on the other. The former are precise beyond anything reasonable; we know virtually nothing about the latter. Let us pause for a minute at this last statement. Are there really no regularities concerning what we just called initial conditions?

The last statement would certainly not be true if the laws of nature (2), (2a) were adopted, *i.e.,* if we considered the $\ddot{r}_i$ as part of the initial conditions. In this case, there would be a relation, in fact the precise relation (1), between the elements of the initial conditions. The question, therefore, can be only: are there any relations between what we really do consider as initial conditions. Formulated in a more constructive way: how can we ascertain that we know all the laws of nature relevant to a set of phenomena? If we do not, we would determine unnecessarily many initial conditions in order to specify the behavior of the object. One way to ascertain this would be to prove that all the initial conditions can be chosen arbitrarily - a procedure which is, however, impossible in the domain of the very large (we cannot change the orbits of the planets), or the very small (we cannot precisely control atomic particles). No other equally unambiguous criterion is known to me, but there is a distinguishing property of the correctly chosen, that is minimal set, of initial conditions which is worth mentioning.

The minimal set of initial conditions not only does not permit any exact relation between its elements, on the contrary, there is reason to contend that these are, as a rule, as random as the externally imposed, gross constraints allow. I wish to illustrate this point first on an example which, at first, seems to contradict it because this example shows its power, and also its weakness, best.

Let us consider for this purpose again our planetary system. It was mentioned before that the approximate regularities in the initial conditions, that is the determinants of the orbits, led Kepler to the considerations which were then left by the wayside by Newton. These regularities form the apparent counterexample to the aforementioned thesis. However, the existence of the regularities in the initial conditions is considered so unsatisfactory that it is felt necessary to show that the regularities are but a consequence of a situation in which there were no regularities. Perhaps v. Weizäcker's attempt in this direction[3] is most interesting: he assumes that, originally, the solar system consisted of a central star, with a gas in rotation, but otherwise in random motion, around it. He then deduces the aforementioned regularities of the planetary system, now called Bode's law, from his assumption. More generally, one tries to deduce almost all «organized motion», even the existence of life, in a similar

fashion. It must be admitted that few of these explanations have been carried out in detail[4] but the fact that such explanations are attempted remains significant.

The preceding paragraph dealt with cases in which there is at least an apparent evidence against the random nature of the uncontrolled initial conditions. It attempted to show that the apparently organized nature of these initial conditions was preceded by a state in which the uncontrolled initial conditions were random. These are, on the whole, exceptional situations. In most cases, there is no reason to question the random nature of the non-controlled, or non-specified, initial conditions and the random nature of these initial conditions is supported by the validity of the conclusions arrived at on the basis of the assumption of randomness. One encounters such situations in the kinetic theory of gases and, more generally, whenever one describes processes in which the entropy increases. Altogether, then, one obtains the impression that whereas the laws of nature codify beautifully simple regularities, the initial conditions exhibit, as far as they are not controlled, equally simple and beautiful irregularity. Hence, there is perhaps little chance that some of the former remain overlooked.

The preceding discussion characterized the laws of nature as regularities in the behavior of an object. In quantum theory, this is natural: the laws of quantum mechanics can be suitably formulated as correlations between subsequent observations on an object. These correlations are the regularities given by the laws of quantum mechanics[5]. The statement of classical theory, its equations of motion, are not customarily viewed as correlations between observations. It is true, however, that their purpose and function is to furnish such correlations and that they are, in essence, nothing but a shorthand expression for such correlations.

Laws of Nature and Invariance

We have ceased to expect from physics an explanation of all events, even of the gross structure of the universe, and we aim only at the discovery of the laws of nature, that is the regularities, of the events. The preceding section gives reason for the hope that the regularities form a sharply defined set, and are clearly separable from what we call initial conditions, in which there is a strong element of randomness. However, we are far from having found that set. In fact, if it is true that there are precise regularities, we have reason to

believe that we know only an infinitesimal fraction of these. The best evidence for this statement derives perhaps from a fact which was mentioned here by Yang six years ago: the multiplicity of the types of interactions. Yang mentioned four of them: gravitational, weak, electromagnetic, and strong, and it now seems that there are two types of strong interactions. All these play a role in every process, but it is hard, if not impossible, to believe that the laws of nature should have such complexity as implied by four or five different types of interactions between which no connection, no analogy, can be discovered.

It is natural, therefore, to ask for a superprinciple which is in a similar relation to the laws of nature as these are to the events. The laws of nature permit us to foresee events on the basis of the knowledge of other events; the principles of invariance should permit us to establish new correlations between events, on the basis of the knowledge of established correlations between events. This is exactly what they do. If it is established that the existence of the events A, B, C,. . . necessarily entails the occurrence of X, then the occurrence of the events A', B', C',. . . also necessarily entails X', if A', B', C',. . . and X' are obtained from A, B, C,. . . and X by one of the invariance transformations. There are three categories of such invariance transformations:

(a) euclidean transformations: the primed events occur at a different location in space, but in the same relation to each other, as the unprimed events.

(b) time displacements: the primed events occur at a different time, but separated by the same time intervals from each other, as the unprimed ones.

(c) uniform motion: the primed events appear to be the same as the unprimed events from the point of view of a uniformly moving coordinate system.

The first two categories of invariance principles were always taken for granted. In fact, it may be argued that laws of nature could not have been recognized if they did not satisfy some elementary invariance principles such as those of Categories (a) and (b) - if they changed from place to place, or if they were also different at different times. The principle (c) is not so natural. In fact, it has often been questioned and it was an accomplishment of extraordinary magnitude, on the part of Einstein, to have reestablished it in his special theory of relativity. However, before discussing this point further, it may be useful to make a few general remarks.

The first remarkable characteristic of the invariance principles which were enumerated is that they are all geometric, at least if four-dimensional space-time is the underlying geometrical space. By this I mean that the invariance transformations do not change the events; they only change their location in space and time, and their state of motion. One could easily imagine a prin-

ciple in which, let us say, protons are replaced by electrons and *vice versa,* velocities by positions, and so on[6].

The second remarkable characteristic of the preceding principles is that they are invariance rather than covariance principles. This means that they postulate the same conclusion for the primed premises as for the unprimed premises. It is quite conceivable that, if certain events A, B, . . . take place, the events X_3, X_2, X_3... will follow with certain probabilities p_1, p_2, p_3. . . From the transformed events A', B', C', the transformed consequences X_1', X_2', X_3'. . . could follow with changed probabilities such as $p_1'=p_1(1-p_1+\sum pn^2), p_2'=p_2(1-p_2+\Sigma p_n^2)$, . . . but this is not the case; we always had $p_i' = p_i$.

These two points are specifically mentioned because there are symmetry principles, the so-called crossing relations[7], which *may be* be precisely valid and which surely do not depend on specific types of interactions. In these regards they are, or may be, similar to the geometric invariance principles. They differ from these because they do change the events and they are covariance rather than invariance principles. Thus, from a full knowledge of the cross section for neutron-proton scattering, they permit one to obtain some of the neutron-antineutron collision cross sections. The former events are surely different from the neutron-antineutron collisions and the cross sections for the latter are not equal to the neutron-proton cross sections but are obtained from these by a rather complicated mathematical procedure. Hence, the crossing relations are not considered to be geometrical symmetry conditions and they will not be considered here. Similarly, we shall not be concerned with the dynamic symmetry principles which are symmetries of specific interactions, such as electromagnetic interactions or strong interactions, and are not formulated in terms of events[7].

On the other hand, it should be noted that the invariance principles themselves depend on the dividing line between initial conditions and laws of nature. Thus, the law of nature (2) or (2a), obtained from Newton's principle by differentiation with respect to time, is invariant also under the transformation to a uniformly accelerated coordinate system

$$\mathfrak{r}_i' = \mathfrak{r}_i + t^2 a \qquad t' = t \tag{3}$$

where a is an arbitrary vector. Naturally, this added principle can have no physical consequence because, if the initial conditions $\mathfrak{r}_i, \dot{\mathfrak{r}}_i, \ddot{\mathfrak{r}}_i$ are realizable *(i.e.,* satisfy (I)), the transformed initial conditions $\mathfrak{r}_i' = \mathfrak{r}_i, \dot{\mathfrak{r}}_i' = \dot{\mathfrak{r}}_i, \ddot{\mathfrak{r}}_i' = \ddot{\mathfrak{r}}_i + 2a$ cannot be realizable.

The symmetry principles of the preceding discussion are those of New-

tonian mechanics or the special theory of relativity. One may well wonder why the much more general, and apparently geometrical, principles of invariance of the general theory have not been discussed. The reason is that this writer believes, in conformity with the views expressed by V. Fock[8], that the curvilinear coordinate transformations of the general theory of relativity are not invariance transformations in the sense considered here. These were so-called active transformations, replacing events A, B, C,. . . by events A', B', C',. . . and unless active transformations are possible, there is no physically meaningful invariance. However, the mere replacement of one curvilinear coordinate system by another is a «redescription» in the sense of Melvin[9]; it does not change the events and does not represent a structure in the laws of nature. This does not mean that the transformations of the general theory of relativity are not useful tools for finding the correct laws of gravitation; they evidently are. However, as I suggested elsewhere[7], the principle which they serve to formulate is different from the geometrical invariance principles considered here.

The Use of Invariance Principles, Approximate Invariances

The preceding two sections emphasized the inherent nature of the invariance principles as being rigorous correlations between those correlations between events which are postulated by the laws of nature. This at once points to the use of the set of invariance principles which is surely most important at present: to be a touchstone for the validity of possible laws of nature. A law of nature can be accepted as valid only if the correlations which it postulates are consistent with the accepted invariance principles.

Incidentally, Einstein's original article which led to his formulation of the special theory of relativity illustrates the preceding point with greatest clarity[10]. He points out in this article that the correlations between events are the same in coordinate systems in uniform motion with respect to each other, even though the causes *attributed* to these correlations at that time did depend on the state of motion of the coordinate system. Similarly, Einstein made the most extensive use of invariance principles to guess the correct form of a law of nature, in this case that of the gravitational law, by postulating that this law conform with the invariance principles which he postulated[11]. Equally remarkable is the present application of invariance principles in quantum electrodynamics. This is not a consistent theory - in fact, not a theory in the proper sense because its equations are in contradiction to each other. However, these

contradictions can be resolved with reasonable uniqueness by postulating that the conclusions conform to the theory of relativity[12]. Another approach, even more fundamental, tries to axiomatize quantum field theories, the invariance principles forming the cornerstone of the axioms[13]. I will not further enlarge on this question because it has been discussed often and eloquently. In fact, I myself spoke about it but a short time ago[7].

To be touchstones for the laws of nature is probably the most important function of invariance principles. It is not the only one. In many cases, consequences of the laws of nature can be derived from the character of the mathematical framework of the theory, together with the postulate that the law - the exact form of which need not be known - conform with invariance principles. The best known example herefor is the derivation of the conservation laws for linear and angular momentum, and for energy, and the motion of the center of mass, either on the basis of the Lagrangian framework of classical mechanics, or the Hilbert space of quantum mechanics, by means of the geometrical invariance principles enumerated before[14]. Incidentally, conservation laws furnish at present the only generally valid correlations between observations with which we are familiar; for those which derive from the geometrical principles of invariance it is clear that their validity transcends that of any special theory - gravitational, electromagnetic, etc. - which are only loosely connected in present-day physics. Again, the connection between invariance principles and conservation laws - which in this context always include the law of the motion of the center of mass - has been discussed in the literature frequently and adequately.

In quantum theory, invariance principles permit even further reaching conclusions than in classical mechanics and, as a matter of fact, my original interest in invariance principles was due to this very fact. The reason for the increased effectiveness of invariance principles in quantum theory is due, essentially, to the linear nature of the underlying Hilbert spacers. As a result, from any two state vectors, Ψ_1 and Ψ_2, an infinity of new state vectors

$$\Psi = a_1\Psi_1 + a_2\Psi_2 \tag{4}$$

can be formed, a_1 and a_2 being arbitrary numbers. Similarly, several, even infinitely many, states can be superimposed with arbitrary coefficients. This possibility of superposing states is by no means natural physically. In particular even if we know how to bring a system into the states Ψ_1 and Ψ_2, we cannot give a prescription how to bring it into a superposition of these states. This prescription would have to depend, naturally, on the coefficients with which

the two states are superimposed and is simply unknown. Hence, the superposition principle is strictly an existence postulate-but a very effective and useful existence postulate.

To illustrate this point, let us note that in classical theory, if a state, such as a planetary orbit, is given, another state, that is another orbit, can be produced by rotating the initial orbit around the center of attraction. This is interesting but has no very surprising consequences. In quantum theory, the same is true. In addition, however, the states obtained from a given one by rotation can be superimposed as a result of the aforementioned principle. If the rotations to which the original state was subjected are uniformly distributed over all directions, and if the states so resulting are superimposed with equal coefficients, the resulting state has necessarily spherical symmetry. This is illustrated in the Fig. 1 in the plane case. This construction of a spherically symmetric state could fail only if the superposition resulted in the null-vector of Hilbert space in which case one would not obtain any state. In such a case, however, other coefficients could be chosen for the superposition-in the plane case the coefficients $e^{im\varphi}$ where φ is the angle of rotation of the original state-and the resulting state, though not spherically symmetric, or in the plane case axially symmetric - would still exhibit simple properties with respect to rotation. This possibility, the construction of states which have either full rotational symmetry, or at least some simple behavior with respect to rotations, is the one which is fundamentally new in quantum theory. It is also conceptually satisfying that simple systems, such as atoms, have states of high symmetry.

The superposition principle also permits the exploitation of reflection sym-

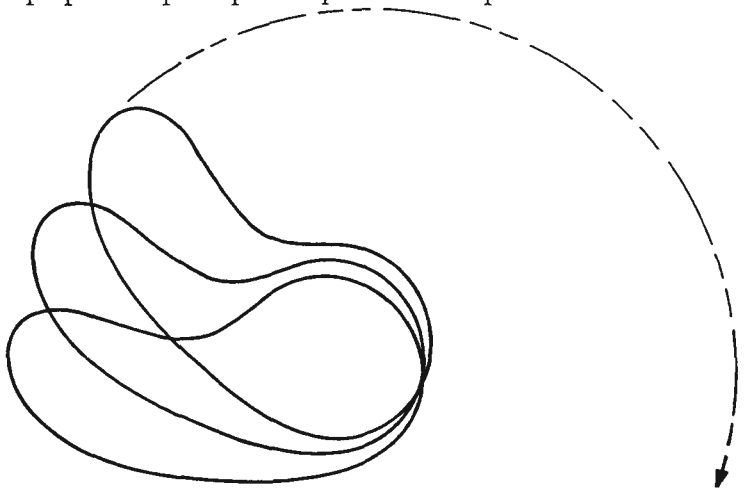

Fig. 1.

metry. In classical mechanics as well as in quantum mechanics, if a state is possible, the mirror image of that state is also possible. However, in classical theory no significant conclusion from this fact is possible. In quantum theory, original state and mirror image can be superimposed, with equal or oppositely equal coefficients. In the first case, the resulting state is symmetric with respect to reflections, in the second case antisymmetric. The great accomplishment of Lee and Yang, which was mentioned earlier[1], was just a very surprising reinterpretation of the physical nature of one of the reflection operations, that of space reflection, with the additional proof that the old interpretation cannot be valid. The consideration of «time inversion» requires rather special care because the corresponding operator is antiunitary. Theoretically, it does lead to a new quantum number and a classification of particles[16] which, however, has not been applied in practice.

My discussion would be far from complete without some reference to approximate invariance relations. As all approximate relations, these may be very accurate under certain conditions but fail significantly in others. The critical conditions may apply to the state of the object, or may specify a type of phenomena. The most important example for the first case is that of low velocities. In this case, the magnetic fields are weak and the direction of the spins does not influence the behavior of the other coordinates. One is led to the Russel-Saunders coupling of spectroscopy[17]. Even more interesting should be the case of very high velocities in which the magnitude of the rest mass becomes unimportant. Unfortunately, this case has not been discussed in full detail even though there are promising beginnings[18].

Perhaps the most important case of special phenomena in which there are more invariance transformations than enumerated before is rather general : it comprises all phenomena, such as collisions between atoms, molecules, and nuclei, in which the weak interaction, which is responsible for beta decay, does not play a role. In all these cases, the parity operation is a valid invariance operation. This applies also in ordinary spectroscopy.

In another interesting special type of phenomena the electromagnetic interaction also plays a subordinate role only. This renders the electric charge on the particles insignificant and the interchange of proton and neutron, or more generally of the members of an isotopic spin multiplet, becomes an invariance operation. These, and the other special cases of increased symmetry, lead to highly interesting questions which are, furthermore, at the center of interest at present. However, the subject has too many ramifications to be discussed in detail at this occasion.

1. Chen Ning Yang, The law of parity conservation and other symmetry laws of physics; Tsung Dao Lee, Weak interactions and nonconservation of parity, *Nobel Lectures, Physics*, 1942-1962, Elsevier, Amsterdam, 1964, pp. 393-403 and pp. 406-418.
2. See, for instance, A. C. Crombie, *Augustine to Galileo*, Falcon Press, London, 1952, pp. 316 ff. The growth of the understanding of the realm of the explainable, from the end of the 13th century on, can be traced through almost every chapter of this book.
3. C.F.V. Weizsäcker, *Z. Astrophys.*, 22(1944) 319; *S.* Chandrasekhar, *Rev. Mod. Phys.*, 18 (1946) 94.
4. An interesting and well understood case is that of « focussing collisions » in which neutrons, having velocities which are rather high but with random orientation, are converted into lower velocity neutrons but with preferential directions of motion. See R.H. Silsbee, *J. Appl. Phys., 28* (1957) 1246; Chr. Lehmann and G. Leibfried, *Z. Physik*, 172 (1963) 465.
5. See, for instance, the section *What is the State Vector* in the writer's article, *Am. J. Phys.*, 31(1963) 6.
6. The possibility of an invariance principle in which velocities are replaced by positions, and conversely, was studied by M. Born, *Nature*, 141 **(193 8) 327;** *Proc. Roy. Soc.* (London), *Ser. A*, 165 (1938) 291, 166 (1938) 552.
7. The crossing relations were established by M.L. Goldberger, *Phys. Rev., 99 (1955)* 979; M. Gell-Mann and M.L. Goldberger, *Phys. Rev.*, 96 (1954) 1433. For further literature, see, for instance M.L. Goldberger and K. M. Watson, *Collision Theory*, Wiley, New York, 1964, chapter 10. The relations of the various types of symmetry principles were considered in two recent articles: *Nuovo Cimento, Suppl.*, in the press, and *Phys. Today* , 17 (1964) 34. See also E. Wigner, *Progr. Theoret. Phys.(Kyoto)*, 11 (1954) 437.
8. V. A. Fock, *The Theory of Space, Time and Gravitation*, Pergamon, Oxford, 1959. The character of the postulate of invariance with respect to general coordinate transformations as a geometrical invariance was questioned already by E. Kretschman, *Ann. Physik*, 53 (1917) 575.
9. M. A. Melvin, *Rev. Mod.Phys.*, 32 (1960) 477.
10. A. Einstein, Zur Elektrodynamik bewegter Körper, *Ann. Physik*, 17 (1905) 891.
11. A. Einstein and S.B. Preuss, *Akad. Wiss., (1915) 778, 779, 844; Ann.Physik, 49* **(1916)** 769. Similar results were obtained almost simultaneously by D. Hilbert, *Nachr. kgl. Ges. Wiss.(Göttingen)*, (1915) 395.
12. J. Schwinger, *Phys. Rev.*, 76 (1949) 790. See also S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, Row, Peterson and Co., New York, 1961, Section 15, where further references can also be found.
13. See A. S. Wightman, Quelques Problèmes Mathématiques de la Théorie Quantique Relativiste, and numerous other articles in *Les Problèmes Mathématiques de la Théorie Quantique des Champs*, Centre National de la Recherche Scientifique, Paris, 1959.
14. G. Hamel, *Z. Math. Physik*, 50 (1904) 1; G. Herzglotz, *Ann. Physik, 36* (1911) 493; F. Engel, *Nachr. Kgl. Ges. Wiss. (Göttingen)*, (1918) 171; E. Noether, *Nachr. Kgl. Ges. Wiss. (Göttingen)*, (1918) 235; E. Bessel-Hagen, *Math.Ann.*, 84 (1921) 258; the quantum theoretical derivation given by E. Wigner, *Nachr. Kgl. Ges. Wiss. (Göttingen)*,

(1927) 375, contains also the parity conservation law which was shown, in ref.1, to be only approximately valid. See also the article of ref.16.

15. I heard this remark, for the first time, from C. N. Yang, at the centennial celebration of Bryn Mawr College.

16. See the writer's article Unitary Representations of the Inhomogeneous Lorentz Group including Reflections, in *Elementary Particle Physics*, Gordon and Breach, New York, 1964, for a systematic discussion of the reflection operations.

17. See the writer's book, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, Vieweg, Braunschweig, 1931 . or the English translation by J. Griffin, Academic Press, New York, 1959 .

18. H.A.Kastrup, *Physics Letters*, 3 (1962) 78. The additional invariance operations probably form the conformal group. This was discovered by E. Cunningham *(Proc. London Math. Soc.*, 8 (1909) 77) and by H.Bateman *(Proc. London Math. Soc.*, 8 (1910) 223) to leave Maxwell's equations for the vacuum invariant, i.e., the equations which describe light, always propagating at light velocity. For more recent considerations, see T. Fulton, F. Rohrlich and L. Witten, *Rev. Mod. Phys., 34* (1962) 442 and Y. Murai, *Progr. Theoret. Phys. (Kyoto),* (1954) 441. The latter articles contain also more extensive references to the subject.