DRAFT - 11/15/86
For Conference Distribution Only

SYMMETRY AND THE LAWS OF NATURE

by

Lloyd Motz
Department of Astronomy
Columbia University
New York, New York

The Fifteenth International Conference on the Unity of the Sciences
Washington, D.C. November 27-30, 1986

Symmetry and the Laws of Nature

Lloyd Motz
Department of Astronomy
Columbia University, N.Y.C.

1. Introduction

When, as intelligent beings, we explore the universe, we probably are most impressed by the great and wonderful variety of objects in it, and our first response to this variety is to arrange these objects into classes, each of which consists of all those that have similar characteristics. Thus we speak of stars, of clouds, of trees, of stones, of mammals, and so on, to take account of the differences (or similarities) that exist among members of each class. The arrangement or classification of objects into broad groups on the basis of similarities seems to be a valid and fruitful procedure, but it is not precise enough when we seek to divide a broad class of, what appear to be similar objects, into subclasses of more intimately related objects. How are we to be guided in doing that? We must, of course, go beyond mere appearances and study group characteristics that may be subsumed under the general heading of symmetries. Here, however, we must be careful, for symmeteries in nature occur at various dimensional levels, and, ultimately, must be explicable in terms of the symmetries associated with the most elementary or basic structures of the universe (molecules, atoms, electrons, nuclei, etc.). A brief consideration of macroscopic symmetries and their relationships to the structures in the universe lead us to a more fruitful and profound way of looking at symmetries and the roles they play in nature.

The symmetry of the snow flake, the gem, the spider's web, and the honey comb, of the bird's nest and the flower, of the earth, the sun and the solar system, all evoke our admiration and wonder. Why and how do such macroscopic symmetries arise? Since these symmetries are the properties of structures, we are naturally led to

MOTZ

the consideration of the symmetries of the forces that govern such structures. I define a "structure" here as an ensemble of particles or bodies (e.g. the solar system) that are united in the same dynamical pattern by a force or forces. From Newton's laws of motion it is clear that all structures originate from forces, for if all the forces in the universe were suddenly eliminated, Newton's first law of motion (the law of inertia) would impose straight line motion on all particles (electrons, proton, etc) and all structures would explode into randomly moving particles. That structures exist with definite symmetries is therefore related to some property of the forces that govern the structures, and since the forces of nature are governed by definite laws, the symmetries (macroscopic and microscopic) must stem from the Natural Laws.

2. Symmetry and the Forces of Nature

Since structures are produced by forces, the observable macroscopic symmetries of structures should be deducible or correspond to the symmetries inherent in the forces. That the symmetries of structures are related to the symmetries of forces is indicated by the way forces beget structures. From Newton's second law of motion we know that a particle, subjected to a force, cannot continue to move in the same straight line with the same speed but must suffer a change in its motion; its speed, its direction of motion, or both must change. A structure is then produced if the force causes the particle to move in a closed path, for then the particle's motion is cyclical and the motion itself or the path of the particle is is a permanet feature and, hence, a structure. If many particles are doing this together under the action of one or more forces, more or less complex structures arise, with symmetries that are related to the forces involved. With this understood we can now trace structural symmetries back to force symmetries.

Physicists have recognized four distinct forces in nature and have listed them, in order of increasing strength, as gravity, the weak force (weak interac-

tion), the electromagnetic force, and the stong force. This classification may be quite misleading, for, whereas gravity, under normal conditions, is, indeed weak, as indicated by our daily experience in lifting objects against the pull of the entire earth, it can, under the proper circumstances (e.g. on the surfaces of black holes) overwhelm the other three forces. Each of these forces has its own distinct characteristics, and they appear to act independently of each other (at least to a first approximation) so that we can study their symmetries independently. A very important general characteristic of these forces is that each is dominant in a particular spatial domain in which its symmetries are impressed on the structures in that domain. Thus gravity's domain is the entire universe and its structures range from comets to galaxies and clusters of galaxies. The weak force, on the other hand, operates over very minute domains whose diameters are of the order of 10^{-16} cm; it monitors radio-activity involving the the emission or absorption of neutrinos and the transformations of nucleons and is dominant in the cores of stars like the sun where the thermonuclear fusion of protons into helium nuclei occurs continuously. Since the first step in this thermonuclear energy generating process is governed by the weak interaction, all life on earth depends on it. The weak interaction makes possible the existence of neutrons, without which nuclei (other than protons) could not exist.

The electromagnetic is the most interesting of the forces, for it is the life force with an incredible variety of simple and complex symmetries. Its domain extends from atomic dimensions (10^{-8} cm), through molecular dimensions and, in a sense, to astronomical dimensions, for the light from the stars, entering our eyes, generates an electromagnetic interaction between the nerves in the retinas of our eyes and the stars. Atoms, molecules (from the simplest like water to the complex like DNA), crystals like diamonds, rocks, liquids, and all living structures are governed by the electromagnetic force. The almost infinite variety of symmetries of such structures can be reduced to or traced back to to the symmetries of the

electromagnetic force.

The nucleus of the atom is the domain of the strong force, whose range is thus about 10^{-13} cm; it can be detected only by particles (e.g. protons and neutrons) that come within that distance of an atomic nucleus. Although the nature of the strong interaction (the nuclear force) is only vaguely understood, the nuclear properties, structures and symmetries stem from the symmetries of this force, which plays its most important role in nature in the build up of heavy nuclei, such as iron, from hydrogen and helium in the very hot interiors of massive stars.

Although at each moment in the history of the universe the four forces operated together, they played more or less important snd dominant roles at different times in the evolution of the universe and in the organization of order and structure from the initial amorphous chaos in which the universe was born: each came upon the stage of the cosmological drama as the dominant actor at the appropriate epoch and then assumed a less dominant role. Thus the universe evolved to its present complexity, symmetry, and order until the conditions in the neighborhoods of stars like the sun permitted the electromagnetic force to organize matter into the highest states of order - living cells.

To elucidate how the symmetries of structures stem from the symmetries of the forces that produce these structures I must first explain the phrase "symmetry of a force" which is most easily done for gravity, the most thoroughly understood of the four forces. We become aware of the symmetry of gravity when we pass from the crawling to the walking stage of our lives and note the difference between the vertical and horizontal orientations of objects; the latter is more stable than the former. We also learn, in time, that freely falling bodies always fall vertically, that is, at right angles to the earth's surface no matter where we may be. This tells us that the earth's gravity acts along a line perpendicular to the earth's surface and hence, since the earth is very nearly a sphere, along a radius. Thus

the earth's gravitational force pulls all things towards its center. This means that the gravitational force is spherically symmetric; that is, it is the same along all directions radiating from the source of the gravitational field. This, however, is true only if the source of the gravitational force is a concentrated bit of matter (that is, a point particle) or a sphere in which the matter is arranged about the center in layers or shells of uniform density. If this is not so, that is, if the source is not a point or a sphere or if the matter is not properly arranged in it if it is a sphere, the spherical symmetry of the force is broken. Thus the earth's gravity is not exactly spherically symmetric because the rotation of the earth has changed its shape from a perfect sphere to an oblate spheroid and has thus broken the spherical symmetry of its gravitational field.

Another important symmetry of the gravitational field relates to the masses of the gravitationally interacting bodies. Newton saw that the gravitational force between 2 massive particles can depend only on their masses and their separation (the distance between them). The symmetry stemming from the dependence of the force on the separation of the two particles is, of course, the spherical symmetry I described above, but the symmetry related to the masses is somewhat more subtle; it states that the force between the two masses must remain the same if the two masses are interchanged. From this we deduce that the formula that describes the strength of the gravitational force must depend on the product of the two masses. It is remarkable that just these two simple symmetries, the spatial spherical symmetry (the force depends only on the distance between the particles and acts along the line connecting them) - and the mass exchange symmetry (the force is proportional to the product of the masses) - describe Newton's law of gravity completely.

We can now relate these two symmetries to the structural and dynamical symmetries of gravitationally bound masses. Consider first a collection of particles

particles moving about randomly, without any overall angular momentum, that is, without any net rotation, like the molecules and grains of dust in a non-rotating nebulous cloud; such gaseous nebulae occur throughout the spiral arms of our galaxy. The particles in this cloud pull upon each other gravitationally, and since these pulls are always along the lines connecting the particles, their net affect is to bring the particles together to form a sphere. Thus the symmetry of the gravitational force leads to spherically symmetric structures such as stars, planets and huge clusters of stars (globular clusters). We thus find a direct relationship between the symmetry of the gravitational force and the spatial symmetry of its structures. This symmetry is broken by rotation as revealed in the dish-like structures of our solar system and galaxies whose symmetries may be defined as cylindrical rather than spherical. The reason rotation breaks the spherical symmetry of the gravity of a point source or a sphere is that rotation introduces an inertial force (the so-called centrifugal force) which is an outward gravity-like force perpendicular to the axis of rotation which becomes an axis of symmetry. The temporal symmetry of gravity is also broken in a strange way in the interior of a black hole. Under ordinary conditions the direction in which a body can move in a gravitational field is independent of the flow of time, but in a black hole space and time are interchanged in such a way that moving out of the black hole is impossible because such motion would mean going from the future into the past.

The symmetries of the other three forces are more complex than those of gravity and more difficult to describe. Thus the symmetry of the electromagnetic force between two unlike charges (for example, between the positively charged proton and the negatively charged electron) at rest, is the same as that of the gravitational field. But this symmetry is broken for a mixture of like and unlike moving charges. Like charges repel each other, which complicates, or breaks, the symmetry, and motion generates magnetic forces which break the symmetry still further, but leave others, that is, electromagnetic and space-time rotations, unbroken. In any case the structural symmetry of an atom, with its electrons revolving around the nucleus, stems from the spherical symmetry of the electrostatic force between the nucleus and the electrons. The symmetry of the strong force is not easy to perceive from its mathematical formulations because no such formulation is known; but that the strong force possesses a pronounced symmetry is clear from the more or less spherical structures

of heavy nuclei, which, to some extent, resemble globular clusters of stars. An interesting symmetry of this force is that it is charge independent; that is, the strong forces between two protons between two neutrons, and between a neutron and a proton are the same. The weak force is characterized by or manifests a certain asymmetry, as described below, rather than any notable symmetry; physicists have discovered that wherever the weak force produces or governs certain events, these events are not mirror symmetric, that is, they are not symmetric with respect to ar interchange of left and right. The weak force, which is characterized by the emission or absorption of neutrinos or anti-neutrinos thus enables us to distinguish between the real world and the mirror image of the real world, as I show later.

2.1 Dynamical Symmetries

Thus far I have been discussing the relationship between the structural symmetries in the universe and the symmetries of the forces that produce these structures; but now I consider much deeper symmetries which are related to the dynamical properties of the structures (or rather, of motions of bodies that constitute the structures such as planets in a solar system, electrons in an atom, etc). These considerations lead us quite naturally to even more profound symmetry relationships - those between conservation principles and space and time. Again these symmetry relationships are best revealed in the dynamics of gravitationally bound structures such as our solar system. During the thirty years preceding Newton's birth Johannes Kepler, using the observational planetary data collected by Tycho Brahe, stated his three famous laws of planetary motion: 1) each planet moves around the sun in its own ellipse one focus of which is occupied by the sun (thus all the planetary ellipses have one focus in common); 2) the line (radius vector) from the sun to a planet sweeps out equal areas in equal times as the planet moves around the sun (the law of areas); 3) the square of the period of a planet (the time it takes it to revolve around the sun) is proportional to the cube of its mean distance from the sun: or, put differently, the square of a planet's period divided by the cube of its mean distance is the same number for all planets. This law, as Kepler stated it, is not quite correct but it is good enough for our discussion.

Two kinds of symmetries are expressed in these laws, which, of course, must be related to each other and, in some way, to the symmetries contained in Newton's law of gravity: the first of these two kinds of symmetries is the geo-

metrical symmetry as expressed in the elliptical orbits of the planets (the first law) and the second kind is the dynamical symmetry as expressed in Kepler's second and third laws. The relationship between these two symmetry categories is revealed in the two geometrical parameters that characterize an ellipse: its size, or semi-major axis, and its shape or eccentricity, and the two dynamical parameters that characterize the planet's motion: its energy, and its angular momentum (rotational motion). The solution of the gravitational two body problem (e.g. a planet revolving around the sun, the Kepler Problem) shows in a very simple and elegant way that the relative orbit of the two bodies is an ellipse, that their total energy (kinetic plus potential energies) is given by the size of the ellipse, its semi-major axis, and that their total angular momentum is given by the shape or eccentricity of the ellipse. Thus the symmetry contained in Newton's law of gravity is translated into the geometrical symmetry of the two body orbit and into the dynamical symmetry of the motion; this leads us to a still more profound insight into symmetries and the laws of nature as delineated below.

3. The Conservation Principles

It is easy to deduce, with little more than elementary algebra, Kepler's three laws of planetary motion from the basic dynamical conservation principles without referring explicitly either to Newton's law of gravity or to his laws of motion. Of course the laws of motion and the law of gravity are contained in the conservations principles so that these tell us no more than do the former, but they give us a deeper insight into the dynamics than do the laws of motion, and each conservation principle is directly related to and deducible from a specific symmetry, so that the role of symmetries in the universe is best revealed in their relationship to the conservation principles.

As the science of mechanics evolved from its rudimentary form, after Newton's discoveries, to its present beautiful mathematical structure (primarily from the work of the late eighteenth and early nineteenth century mathematical physicists such as Lagrange, Euler, Laplace, Poisson, Gauss, Hamilton, Jacobi) it became clear that the dynamical evolution of a system (aggregate of particles) could be deduced from very general minimal principles, subject to certain definite constraints. The constraints are the conservation principles and the minimal principles state that, as the system evolves, it must change in such a way that a certain physical quantity associated with the system changes by the smallest possible amount.

A famous example of a minimal principle is Fermat's principle of least time, which states that a ray of light, moving from an initial to a final point, moves along that path in which it spends the least time. Another equally famous and extremely productive minimal principle was first introduced into Newtonian mechanics by the eighteenth century mathematical physicist Maupertuis and later generalized and extended by Hamilton: it states that a particle, subject to forces, moves in such a manner that a certain quantity called its action associated with its motion, changes by the smallest possible amount along the particle's orbit. This is the famous principle of least action (action, in its simplest form, is the momentum of a particle times its change in position). Since these principles themselves are subject to the conservation principles, I now discuss the latter, which originated from Newtonian mechanics, but which have been extended considerabley since then.

We know from our daily experience that a continual recyclying of matter goes on in nature so that the concept of conservation is not alien to us - things change from one form to another but the total number of the elementary particles which constitute matter, whatever they may be, remains the same. This is the essence of classical atomic theory which, in chemistry, is called the conservation of mass; as we know, it is only an approximate conservation principle. The conservation principles that govern classical dynamics are of a non-material nature and can be deduced from the laws of motion; since they deal with motion rather than with matter, they are more difficult to perceive than the conservation of mass. First we have the conservation of momentum (the momentum of a particle is defined as the product of the particle's mass and its velocity), which states that if no net external forces act on an aggregate of interacting particles, the total momentum of the aggregate must remain constant, regardless of how the constituent particles move about or interact with each other. The total momentum of an aggregate of particles is obtained by summing the momenta of the individual particles (a vector sum, since momentum is a vector); the momentum of any particular particle varies from moment to moment, but the net effect of the variations of all the particles is zero. This conservation principle, which stems from Newton's law of action and reaction, is extremely important in analyzing the interactions between colliding particles and the results produced by such collisions, and the decay of particles such as neutrons. The conservation of momentum is related to a spatial symmetry, as we shall see in the next section, because it defines an important point in an aggretate of particles called the center of mass, which permits an alternate way of stating the principle: if no net external force acts on a system of particles, the center of mass of this

system remains at rest (if it was initially at rest) or continues to move with the same speed in the same straight line.

Closely associated with the conservation of momentum is the conservation of energy, which, again, stems from Newton's laws of motion and his law of gravity. This conservation principle was first thought to be valid only for the mechanical energy (kinetic plus potential) of bodies moving without friction, but it was later extended to motion with friction when heat was recognized as a form of energy.* It was then called the first law of thermodynamics. We now know, from the theory of relativity, that mass and energy are equivalent so that the conservation of energy includes the conservations of mass; mass by itself is not conserved.

Taking relativity into account, which has 4-dimensional space-time rotational symmetry, we must combine the conservation of energy (which includes mass) and the conservation of momentum into a single conservation principle. Indeed, the theory of relativity tells us that energy and momentum of a system are not conserved separately for all observers, but that a single quantity, called the energy-momentum 4-vector, is conserved. In the calculation of the momentum and energy of a system, the energy and the momentum of each photon must be included and the energy corresponding to the mass of each particle must also be taken into account. The conservation of energy simply ensures that no process will occur if not enough energy is available for the process. If enough energy is available, a process will occur spontaneously, unless some other conservation principles prohibit it. Since the total energy of the system (the total measured mass times the square of the speed of light plus the energy of each photon) before the spontaneous process occurs is the same as after the process, what do we mean by "enough energy" available? Why should a process go in one direction rather than in the reverse direction if the total energy must be the same at each step of the process? To answer this question we must consider the masses alone. If the total mass of a system is larger than its total mass after the process, the process will occur spontaneously. In other words, the "energy available" for a spontaneous process is the difference between the total mass (the sum of the masses of all the particles in the system) of the system before and after the process. Thus, since the mass of the neutron is somewhat larger than the mass of a proton plus an electron, the neutron spontaneously decays into a proton and an electron. The total energy after the decay is the same as before, but not all of it appears in the form of mass. Some of the original energy appears as kinetic energy or

*See paper by B. Elbeck in this committee

the newly formed particles: the proton, electron, and neutrino.

In addition to conservation of momentum and energy, we also have conservation of angular momentum or rotational motion. The total rotational motion of an isolated system before and after a process must be equal. Rotational motion, in general, consists of two parts: one part arising from the orbital motions of the particles in the system, and the other from the spin of each particle. The total rotational motion or angular momentum is obtained by summing the orbital and spinning motions for all the particles. In discussing the properties and behavior of elementary particles, we do not deal with orbital motions; the only thing that concerns us here is the spin of each particle. The conservation of angular momentum then tells us that, in any process involving the transformation of one group of "elementary" particles into another, the total spin (the spins of particles added together) before and after the process must be the same.

Here we must be careful because spin is a vector (that is, a directed quantity) and adding such quantities is unlike the usual process of additon. We overcome this difficulty by always considering the components of the spin in a particular direction and adding them together. The spins of particles occur in integer and half-integer multiples of a basic unit, which is Planck's constant h divided by 2π, and is written as $\hbar$. In terms of this unit, the spin of both the electron and the proton is $\frac{1}{2}$ and the spin of the photon is 1. Like the electron and the proton, the neutrino also has a spin of $\frac{1}{2}$.

Are there any other conservation principles that must be taken into account in our analysis of the fundamental particles of which matter is constructed? There are a few other important ones. One of these deals with electric charge. The total charge in the universe is constant, at least as far as all evidence indicates; charge can neither be created nor destroyed. Therefore, the total charge of a system must be the same before and after a process occurs. This is the principle of conservation of electric charge. If a new, positively charged particle suddenly appears during a process, a new, negatively charged particle must appear at the same time to compensate for the positive charge.

Charge occurs in integer multiples (positive and negative) of a basic unit, which is the charge on the proton. As far as is known, only three values of this multiple of the unit of charge occur on fundamental particles: -1, 0, +1.

Examples are the electron (-1), the neutrino (0), and the proton (+1). The charge on the photon is also 0. This is called the quantization of electric charge, the cause of which is not known.

This last statement is correct if we consider only those particles that we can observe directly, but if we accept the strong experimental evidence for the composite nature of the proton and neutron, then elementary particles (quarks) exist with electrical charges of -1/3 and +2/3 of the "unit charge"; the concept of a unit charge then has no meaning, for there is no more reason to call the charge on the electron the unit charge than there is to call the charge on either of the two different quarks the unit. In any case the conservation of electric charge applies whether we accept the existence of quarks or not.

The conservation of electric charge enjoyed a dramatic expansion in 1932 with the discovery of the positron, or the anti-electron, whose existence the Dirac relativistic theory of the electron had already predicted. This discovery and Dirac's theory show that electric charges can arise from the vacuum, provided they arise in equal and opposite pairs (particle and anti-particle) so that charge conservation is related to a remarkable symmetry in nature which I shall discuss later.

Another conservation principle deals with the total number of heavy particles or nucleons (protons and neutrons) in the universe. Since no experimental evidence has ever been adduced for the destruction or creation of a heavy particle, we must assume that the total number of heavy particles (protons plus neutrons or any other particles that finally become protons or neutrons) in the universe is conserved. This number must therefore be the same before and after any process. This is the principle of the conservation of baryons, which is basic in particle physics.

Thus, a neutron decays into a proton and two light particles (electron and neutrino) so that we start with one heavy particle (neutron) and end with another one (proton). Light particles like electrons, and neutrinos are not conserved individually, as is clear from the β-decay process (e.g., the decay of the neutron). Note, however, that the total lepton number, where by leptons we mean electrons, neutrinos, and muons (and their antiparticle counterparts) is conserved. This follows because a lepton always appears or disappears with an antilepton, so that the total number, counting antileptons as negative, remains constant. This is the principle of the conservation of leptons.

Another important conservation principle, the conservation of parity, entered physics with the discovery of the wave properties of particles, e.g. electrons, which, we now know, behave both like particles and waves, depending upon how we observe them. By conservation of parity, we refer to the behavior of the wave properties of a particle. Since every particle is described by a wave amplitude, which depends on (that is, is a function of) the coordinates of the particle, we can classify particles according to how their wave amplitudes or "wave functions" behave when the coordinates of the particles are replaced by their negative values. This simply means comparing the behavior of a particle in the real world with its behavior as seen through a mirror. If such a reflection leaves the wave amplitude unchanged, we say that the particle has even parity, but if the wave amplitude changes its sign when viewed in a mirror, we say that the parity is odd. Thus, parity can have only two values: +1 (even parity) and -1 (odd parity). Conservation of parity means that the total parity of a system (obtained by multiplying together the parities of the individual particles in the system) is the same before and after a process occurs. As was first predicted by Lee and Yang in 1957, this principle is violated in weak interactions, that is, interactions or processes in which neutrinos are either emitted or absorbed. This is an example of symmetry breaking, which I discuss in detail later; it is astonishing, for one expects the universe to be completely symmetric with respect to the interchange of left and right. This means that, contrary to expectations, the laws that govern the real universe are slightly different from those that govern a mirror image of the real universe; the existence of the neutrino destroys the right-left symmetry.

Another important type of symmetry, associated with the wave properties of particles, but not related to any apparent conservation principle, was discovered in the analysis of the spectral lines of atoms; it stemmed from the introduction of quantum members to describe the dynamics of an electron inside an atom, as required by the quantum theory. Just as classical dynamical theory of planets associates the parameters of a planet's orbit (size, shape, etc) with the planet's dynamical parameters (energy, angular momentum, etc.), as previously described, so quantum theory associates a set of discrete integers, called quantum numbers, with the dynamical parameters of an electron in an atom; the existence of such quantum numbers in quantum theory is equivalent to the conservation principles in classical theory. The motion of an electron (or its

quantum state) is completely defined by its set of quantum numbers (exactly four in number for each electron). The symmetry principle associated with an electron's quantum numbers states that electrons in an atom must so arrange themselves that no two of them have the same set of quantum numbers; named after its discoverer this principle is known as the Pauli enclusion principle.

Its importance cannot be overestimated, for, without it, the existence of the chemical elements and their properties could not be explained - indeed, chemical elements could not exist and one could not explain the periodic table of chemical elements if there were no Pauli exclusion principle. As we know, the chemical properties of groups of elements (e.g. lithium, sodium, potassium and helium, neon, argon) whose atomic numbers (positions in the periodic table) differ by definite integers, are similar; this similarity stems from the Pauli exclusion principle which arranges the electrons in an atom in such a way that the outermost electrons, which determine the atom's chemical properties, are always in the same (or similar) dynamical pattern for any two elements whose atomic numbers (total numbers of electrons) differ but which have the same number of outer electrons.

The full symmetry implications of the Pauli exclusion principle are best revealed in the relationship of the wave properties of particles to the statistics that govern these particles. Statistical mechanics, which grew out of the kinetic theory of gases, is a powerful technique for deducing the gross properties of ensembles of particles from the average or statistical behavior of the individual particles, but classical statistical mechanics had to be replaced by quantum statistical mechanics to take into account the wave properties of particles and the indistinguishability of two or more identical particles. It was then discovered that particles such as electrons, protons, neutrons, and neutrinos, called fermions, which have a half unit of spin (the unit is Planck's constant of the action h divided by 2π) obey one kind of statistics - the Fermi - Dirac statistics - and particles such as pions (mesons) and photons, called bosons, which have zero spin or one unit of spin, obey the Bose-Einstein statistics. The wave function that describes an ensemble of identical fermions (e.g. electrons) must be anti-symmetric with respect to the interchange of any two particles - that is, it must change sign on such

Morz

an interchange (from positive to negative or vice versa), which means that the Pauli exclusion principle applies. The wave function of an ensemble of bosons, on the other hand, is symmetric with respect to the interchange of two identical particles so that the Pauli exclusion principle does not apply to these.

4. Symmetry and The Conservation Principles

The most interesting aspect of symmetry in nature is revealed in its relationship to the conservation principles; today we know that each of these principles stems from some kind of space-time or other type of symmetry. Conservation of momentum, the simplest and most easily perceived of the conservation principles, is related to a simple and easily understood space-time symmetry; namely that the laws of nature are invariant to a shift of one's coordinate system from one point of space-time to any other. The reason for this relationship of momentum conservation and the invariance of the laws of nature to a translation of coordinates is that the total momentum of an ensemble of particles establishes a center of mass of the ensemble which remains unaltered if the entire ensemble is shifted in space. Thus to an observer at the center of mass of the ensemble the total momentum of the ensemble is zero and remains so if the center of mass moves with constant speed in a straignt line, as required by the absence of external forces.

Since, as we have seen, conservation of energy is closely associated with conservation of momentum, the symmetry to which it (energy conservation) is related or from which it stems should be closely associated with the spatial symmetry of momentum, and it is; if the laws of nature are symmetric in time (the same whether time flows forward or backward, and the same at all times) then energy is conserved. In producing the theory of relativity Einstein merged the spatial and temporal symmetries into a single space-time symmetry and hence the two conservation principles into a single momentum-energy conservation principle. The space-time symmetry associated with the theory of relativity leads to the concept of anti-matter (e.g. the positron, the anti-proton, etc) for it suggests, in fact, requires, that just as the world line (the 4-dimensional space-time orbit) of an ordinary particle like the electron is directed from the past to the future, the world line of/an anti-particle may be described as that of a particle directed from the future to the past.

Since angular momentum is to the rotation of a system of bodies as momentum is to its translation, the conservation of angular momentum implies that the laws of nature are invariant to a spatial rotation of the observer's frame of reference (rotational symmetry of the laws of nature).

In his general theory of relativity Einstein enlarged, or generalized, space-time symmetry by stating that the laws of nature must be invariant to any transformation of coordinates; that is, no special coordinate systems are favored over any other by nature. This means that the laws of nature must be formulated in mathematical forms that remain unaltered on transformation from one coordinate system to any other; these forms are tensors which have the very desirable property that they are the same in all coordinate system so that a law expressed in tensor form is automatically correct because it is invariant to coordinate transformations owing to its tensor character. Einstein's field equations of gravity, expressed, as they are, in tensor form, thus relate the symmetry of the gravitational force to the non-euclidean symmetry of space-time in the presence of masses.

To what particular symmetry or invariance principle is the conservation of charge related? This invariance is related to a quantity (or physical entity) called gauge, which already appeared in Maxwell's equations of the electromagnetic field. These equations describe how electric and magnetic fields (field intensities or field strengths) are interrelated and combine to form electromagnetic waves. These equations, which show the remarkable symmetry that exists between electricity and magnetism (a phenomenalogical separation of the electromagnetic field) can be expressed in tensor form so that they possess the invariance demanded by the theory of relativity. The gauge concept stems from the use of potentials, rather than field strengths, to define the electromagnetic field, which simplifies things considerably when one calculates the interaction energy of an electric charge with an electromagnetic field. This interaction energy is then the product of the magnitude of the charge and the magnitude of the electromagnetic potential at the position of the charge.

Since the electromagnetic field strengths can be expressed in terms of the potentials, one can write Maxwell's equations in terms of the potentials but the results are rather complicated. However we have a certain freedom in choosing the potentials; we may add to each of the four potentials the relativ-

istic gradient of an arbitrary function of space-time. This is called a gauge transformation because it alters the scale of the potential; it does not alter Maxwell's equations, however, and so may be chosen in such a way as to simplify the formalism.

The full significance of this gauge invariance for electric charge conservation is revealed in the wave equation for a charged particle interacting with an electromagnetic field. The wave function describing such a particle can be multiplied by an arbitrary complex function of absolute magnitude one without changing the physics of the interaction of the charge and field; this factor, which is written as an exponential and is called the phase of the wave, has no effect on the physics described by the wave function. However, the introduction of such an arbitrary phase factor alters the Schrödinger wave equation because it introduces an additive term in the energy of the charge, which, if uncompensated, would be equivalent to a change in the magnitude of the charge. But this term can be exactly compensated for by an appropriate gauge transformation of the potentials so that charge is conserved.

During the past thirty years many, very short lived, massive particles (hadrons) were discovered which can be arranged into two families of multiplets - baryons (particles of spin 1/2, 3/2, 5/2, etc: that is, fermions) - and mesons (particles of spin 0,1,2, etc; that is, bosons). Within each of these families the particles arrange themselves into smaller groups or multiplets which exhibit very definite symmetries, such as spin, mass, and charge; these have been explained by the introduction of more basic units of matter called quarks. Four different quarks called "up", "down", "strange", and "charm" have been introduced to account for all the hadron multiplets that have been observed, with the assumption that each baryon consists of three quarks and each meson of a quark and an antiquark. A special kind of mathematical symmetry, called SU(3), which refers to a group of transformations has been introduced to describe hadron families. The presence of such a symmetry is equivalent to the conservation of baryon number. The simplest representation of the SU(3) group is a set of eight third order (3 rows and 3 columns) unitary matrices which correspond to the members of the baryon octet.

5. Symmetry Breaking

In the previous section I stressed the relationship between symmetry in nature (that is, the invariance of the laws of nature to certain symmetrical transformations of one's frame of reference) and conservation principles. To be universally valid a conservation principle must hold at all points of space-time under all circumstances, and this, of course, can never be proved empirically. In a sense, then, conservation principles must be taken on faith, but not entirely since the relationship of these principles to symmetries enables us to use symmetries as a guide to correct global conservation principles to replace those that are found not to be universally valid. If an exception to a particular conservation principle is discovered, the symmetry to which it is related is said to be broken, and one then seeks some, as yet, undiscovered force as the cause of such symmetry breaking. Such a force, if present, then replaces the broken symmetry by an enlarged unbroken symmetry. A few examples of this will illustrate this point.

Going back to the Newtonian law of gravity we note that its spherical symmetry for a spherical body is broken if the sphere is rotating and this, we know, stems from an inertial force-the so-called centrifugal force. By combining the gravitational and centrifugal forces into a single space-time curvature Einstein established a higher symmetry than that contained in Newton's law of gravity. Similarly, the apparent violation(symmetry breaking) of the classical principle of conservation of energy manifested by the luminosities of the stars was eliminated by Einstein's replacement of the classical principle by an enlarged energy-momentum-mass conservation principle, which stems from space-time symmetry.

The remarkable relationship of symmetry breaking to the presence of a force that breaks the symmetry is strikingly shown during a change of phase from one state of matter to another. Thus the perfect gas law (essentially a consequence of Newton's law of motion) assumes the absence of forces among the constituent molecules of the gas, but a change of phase of the molcular ensemble from its gaseous state (a state of perfect symmetry) to its liquid state (a state of broken symmetry) indicates molecular forces (the van der Waals forces). A further change of phase from the liquid to the solid (or crystalline) state breaks the symmetry still more with the appearance of homopolar bonded molecules, which can be produced only by the quantum mechanical exchange force; this kind of force has no classical counterpart.

Other types of unexplained broken symmetries occur in the universe on a macroscopic scale. Thus time flows in only one direction; entropy always increases; the universe is expanding and not contracting; the number of photons in the universe exceeds the number of protons by a factor of 10^{10}, and so on. On a microscopic scale one of the most important examples of symmetry breaking is that of the interchange of right and left handedness (nonconservation of parity) as discovered by Lee and Yang. Classical physics states that the laws of physics are the same, (or should be the same) when expressed in a left handed frame of reference as when expressed in a right handed frame. (A mirror image universe must be governed by the same laws as the real universe). Lee and Yang found that this does not hold for phenomena that involve neutrinos.

The reason that neutrinos break reflection symmetry is that they have left handedness in the following sense: an observer from whom a neutrino is receding always sees the neutrino spinning counter-clockwise (like a left-handed screw being screwed into a wall). As seen in a mirror, however, such a receding neutrino, still spinning counter-clockwise, would appear to approach the observer, which is contrary to the way an approaching neutrino must behave. Thus the mirror image of our universe gives a wrong description of neutrinos, so that reflection symmetry does not hold universally. This implies the presence of a force, now called the weak interaction, which is said to break the reflection or mirror image symmetry. The weak interaction accounts for the beta-decay of the free neutron and of nuclei which have too large a neutron-proton imbalance to be stable.

That parity is not conserved (that is, that reflection symmetry is broken) when neutrinos are involved points to a more general or a higher type of symmetry which involves antiparticles, as well as particles, and <u>time reversal</u>, that is, the flow of time from the future to the past as well as from the past to the future. If these three symmetries, each of which is broken by itself, is combined into a single <u>supersymmetry</u>, this supersymmetry is conserved. We may imagine this as the reflection of our universe in a supermirror which changes left to right (P), changes particle to anti-particle (C), and reverses the flow of time (T). The laws that govern the universe as we see it also govern the PCT, that is, the reflected, universe.

Another important global symmetry that is broken (at least as far as our observations until now indicate) is that of matter and anti-matter; the equations (laws) of relativity and quantum mechanics show no preference for matter over anti-matter and yet very little (if any) anti-matter has been found. To explain this asymmetry particle physicists have proposed that when the universe was very young (about 10^{-35} sec old) and both particles and anti-particles were equally abundant, a special kind of field of force was present which shifted the balance very slightly in favor of particles so that when the universe cooled off slightly, most of the particles, (all but about one in every ten billion), were annihilated by all the anti-particles, so that only particles and billions of photons per particle (the present cosmic radiation) were left over. No evidence whatsoever for such a force has ever been adduced, however, but it is argued that such a force must exist because the laws of nature, as indicated by the behavior of K-mesons, are not entirely invariant to time reversal, owing to this force. However, that this strange behavior of K-mesons can be explained in some other way has not been ruled out.

The most important application recently of the concept of symmetry breaking in particle physics has been to the development of the theory of the electro-weak force, which purports to be the unification of the electromagnetic and the weak forces. Such an attempted unification encounters the following difficulties: 1.) the electromagnetic force is much stronger than the weak force, and 2.) the carrier of the electromagnetic force is the photon, a massless chargeless particle, whereas the weak force is carried (or said to be carried) by three very massive particles (the intermediate bosons), two of which, $\underline{\underline{W}}^+$,$\underline{\underline{W}}^-$, are charged and one of which, $\underline{\underline{Z}}^o$, is neutral. These difficulties did not deter or hinder the electro-weak theorists who argue that the present asymmetry in the electromagnetic and weak forces stems from symmetry breaking that occurred in the first 10^{-35} sec. of the universe's life when the temperature was millions of trillions of degrees. All the forces (except gravity, which is left entirely out of account) they say were then equally strong, and the photon, the two $\underline{\underline{W}}$'s, and the $\underline{\underline{Z}}$ were massless. Thus complete symmetry prevailed. But then, as the universe cooled, certain fields, the so-called Higgs fields, (also known as gauge fields) that were present from the very beginning, broke the perfect symmetry by assigning masses to the $\underline{\underline{W}}$'s and $\underline{\underline{Z}}$ bosons but leaving the photons massless. The quanta of the Higgs fields possess mass which they are pictured

as assigning to the W's and the Z by attaching themselves to the latter.

The use of unobserved gauge fields as symmetry breaking devices to account for the observed differences among the known forces has propelled physics into strange, unphysical, mystical byways. A new kind of scholasticism seems to have enveloped particle physicists, who argue, not about "how many angels can dance on the head of a pin," but how many different kinds of gauge fields can "dance" inside a baryon. Particle physicists, in general, have thus beguiled themselves that they have achieved a unification of the forces (gravity excepted); they thus speak of the "Grand Unification Theories" (GUTS), even though, as these gauge theories stand now, they are burdened with 27 arbitrary parameters. To speak of this as a unification of the forces of nature is somewhat arrogant, to say the least, and lacking in the kind of skepticism that has been so fruitful in science in the past.

In a sense the situation in particle physics now is similar to that which prevailed in pre-Copernican cosmology, when each new correction in the observed motions of the planets was explained by the introduction of a new epicycle. Today difficulties in particle physics and their carry over to cosmology and other related phenomena are eliminated (or "swept under the rug") by introducing supersymmetries of all kinds. Thus particle physicsts today talk quite seriously about an overall supersymmetry which introduces a super-space to supplement ordinary four-dimensional space-time. In this super-space, pictured as a spinning space with non-commuting coordinates, fermions and bosons are interchanged, so that the force of gravity takes on a special form called super-gravity, which, it is hoped, will unify gravity with the other three forces. But such unification attempts have failed. Nevertheless, these attempts are still being pursued with the introduction of superstrings (structures with length, but no width) to replace particles as the basic constitutents of all matter, and the introduction of the concept of compactification of dimensions to account for the disappearance of all but four of the many dimensions which supersymmetry and supergravity require. But of all proposals stemming from these concepts the most bizarre is that of "shadow matter," which promulgates the existence of a "shadow" universe whose matter can interact only gravitationally with ordinary matter. The ordinary matter

and the shadow matter in this theory, are pictured as intermixed and existing together in the universe. Since there is no observational evidence, whatsoever, for shadow matter, the advocates of this theory can endow it with whatever physical properties they wish to account for unexplained phenomena.

But this, too, has led to a dead end, and so, on the whole, the introduction of special gauge fields, supersymmetries, broken symmetries, superstrings, compactification, etc. has brought us no closer to a unification of the forces of nature than we were in Einstein's time. It is clear that so long as gravity is left out of unification theories, such theories must fail or, at best, remain inadequate, for gravity, the primary force in the universe, cannot be neglected in the domain of the quarks.