Emmy Noether

Symmetry in Physics: Introduction

Symmetry is a crucial concept in mathematics, chemistry, and biology. Its definition is also applicable to art, music, architecture and the innumerable patterns designed by nature, in both animate and inanimate forms. In modern physics, however, symmetry may be the most crucial concept of all. Fundamental symmetry principles dictate the basic laws of physics, control stucture of matter, and define the fundamental forces in nature.

Some of the most famous mathematicians and physicists had this to say about symmetry:

"I aim at two things: On the one hand to clarify, step by step, the philosophic-mathematical significance of the idea of symmetry and, on the other, to display the great variety of applications of symmetry in the arts, in inorganic and organic nature."

Hermann Weyl (in his book "Symmetry")

"Special relativity emphasizes, in fact is built on, Lorentz symmetry or Lorentz invariance, which is one

of the most crucial concepts in 20th Century Physics."

C. N. Yang (Nobel Laureate in Physics)

"*Look at the symmetry of the laws, i.e., look at the way the laws can be transformed, and leave their form unchanged....*" and, " *Symmetry is fascinating to the human mind; everyone likes objects of patterns that are in some way symmetrical.... but we are most interested in the symmetries that exist in the basic laws themselves.*"

Richard P. Feynman (Nobel Laureate in Physics; in his "Lectures on Physics")

"*I heave the basketball; I know it sails in a parabola, exhibiting perfect symmetry, which is interrupted by the basket. Its funny, but it is always interrupted by the basket.*"

Michael Jordan (former Chicago Bull)

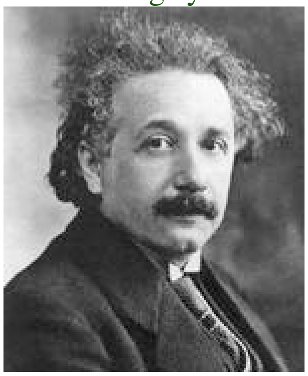
The most powerful microscopes humans have built are the great particle accelerators, such as the Tevatron at Fermilab, in Batavia, Illinois. The Tevatron accelerates protons and antiprotons in opposite directions in a great circle, to energies of one trillion electron volts (as though you had a one trillion volt battery hooked up to a vacuum tube). These particles then collide head-on. The quarks and antiquarks, inside of the protons and anti-protons, themselves collide. By reconstructing the debris from a collision of this kind physicists get a kind-of "photograph" of the structure of matter at the shortest distance scales ever seen. These distances are as small in comparison to an atom as the atom is small in comparison to Michael Jordan's basketball.

By studying physics at these tiny distance scales we can see that the forces of nature begin to share a common property, which is unseen at lower "magnification," at the larger distant scales. Today we understand that all of the fundamental forces in nature are unified, in a sense, by one elegant symmetry principle. This principle is subtle, and therefore it has a fancy name: it is called "local gauge invariance." Eventually we'll try to explain "local gauge invariance" to a wide audience, perhaps in this website, but for now please accept this as a statement of fact.

The discovery of this unifying symmetry principle has allowed us to leap conceptually to distance scales one-thousand-trillion times smaller than can be seen with our most powerful microscopes (particle accelerators). This has allowed us to conceive of what the Universe was like in the first one billionth of one billionth of one billionth of one billionth of a second! At such short distances quantum gravity is active and forbids our normal notions of space and time. There we must use the symmetry principles (and related topological ideas) to imagine theoretically the complete unification of all forces. This leads to a new ideas, to something called the "superstring", and an arcane mathematical system called M-theory that no one yet understands (we really don't even know what "M" stands for). Nevertheless, this is, perhaps, the most symmetry-pregnant logical system ever conceived by the human mind.

Indeed, we revere the fundamental symmetries of nature and we have come to intimately appreciate their subtle consequences. For example, we learn from Emmy Noether that to succumb to a crack-pot's invention requiring us to give up the law of energy conservation would be to give up the notion that time flows symmetrically, with no change in the laws of physics. To give up the notion that the speed of light is a fundamental limitation on the propagation of signals would be to give up the fundamental symmetry of Special Relativity, i.e., Lorentz invariance, and its consequences, such as the equivalence of matter and anti-matter, etc. Symmetry controls physics in a most profound way, and this was the ultimate lesson of the 20th century.

Yet, a sampling of the crucial role of symmetry in physics for the beginning students is at odds with the practice of the complete omission of this beautiful and fundamental topic. Not only is it missing in the high school curriculum, but also in the standard first year college calculus-based physics course. It does not appear in the Standards. The profound relevance of symmetry principles in our understanding of nature is largely a 20th century revelation, beginning with Einstein's view of physics.



The absence of symmetry discourse in our teaching of physics today represents a throw-back to a nineteenth century perspective which seems to permeate the curriculum.

We believe that students are attracted to and motivated by the modern and "sexy", highly visible end products of modern physics, e.g., semiconductors, lasers, nuclear and atomic processes, superconductors, superfluids, Bose-Einstein condensation, the formation of galaxies and black holes, the Big-Bang, quarks and strings. The process of learning about these things in detail takes some six to eight years of undergraduate and graduate physics courses. Only then, if the student chooses a very abstract field of specialization, such as theoretical elementary particle physics, will she begin to see the fundamental role of symmetry in the basic laws of physics. Indeed, even today many practicing physicists have no idea about the concept of, e.g., local gauge invariance!!!

It is possible, nonetheless, to incorporate some of the introductory underlying ideas of symmetry and its relationship to nature into the beginning courses in physics (and mathematics), at the High School and early college level. It has worked successfully for us in countless "experiments," and it is a lot of fun for everyone. These ideas at the outset are really not that difficult. When the elementary courses are spiced with these symmetry ideas, they become enhanced and begin to take on some of the dimensions of a humanities or fine arts study: *Symmetry is one of the most beautiful concepts, and its expression in nature is perhaps the most stunning aspect of our physical world*. We believe that symmetry will prove to be a vehicle for maintaining and enhancing the student's interest in physics at the outset and connecting to the deeper aspects of our relationship and understanding of the physical world.

What follows is a description of a high school or early-college level module that introduces the key ideas of symmetry which, in many examples, tie physics to astrophysics, biology and chemistry. This involves no calculus, but a basic knowledge of geometry and high school junior algebra (with complex numbers, trig, etc.) is prerequisite. This also attempts to reveal some of the modern thinking in a conversational way. We are experimenting in the classroom, in Saturday Morning Physics Lectures at Fermilab, and elsewhere, in the implementation of this approach.

We will continually update this website as our educational experiment in Symmetry proceeds (indeed, as you can see, it isn't finished yet). And don't hesitate to send us suggestions, comments, and even kindly worded complaints.

Emmy Noether

Symmetry in Physics: What is Symmetry?

When a group of students is asked to define "symmetry" the answers they give are generally all correct. For example, to the question "What is Symmetry?" we hear some of the following:

"its like when the sides of an equilateral triangle are all the same, or when the angles are all the same..."

"things are in the same proportion to each other... "

"things that look the same when you see them from different points of view ... "

"its the same thing before and after you do something to it... "

A system is said to possess a symmetry if one can make a change in the system such that, after the change, the system appears exactly the same as before.

Symmetry begins as a visceral human concept. We can see it with our eyes, and hear it with our ears, and feel it with the right sides of our brains. It is often equated to perfection. The ancient architects embodied symmetry into their designs. An ancient Greek temple, a Pharoah's tomb, medieval cathedral, all attempt to reflect the kind of home that a God would choose to live in. The anatomical features of living organisms embody symmetries. The sun and moon appear as perfect spheres. Thus symmetry is invoked to viscerally touch and nurture our needs for a divinity, a perfect order and harmony. Through symmetry we feel something of a spiritual world, beyond our limited and frail human plight. The arts and music have turned in the past century upon thematic issues involving symmetry, indeed often adopting "anti-symmetry" as a thematic element (e.g., the elimination of the tone-center in 12-tonal musical is a movement away from a center of reflection symmetry, yet it remarkably moves toward the symmetry of "translational invariance" which we see in space and time).

Yet, it is the remarkable achievement of the past century that symmetry has become understood as the fundamental ingredient to the formulation of the laws of physics.

Reflections:

In the photograph to the right we see the interior view of the majestic gothic cathedral at Amiens, France. We can use the photograph of the cathedral to illustrate the concept of one kind of symmetry, known as reflection. In the second photograph we have produced a mirror image of the original by reflecting each point on the right into a point of the left. This is the picture the camera would have taken if instead of viewing the cathedral directly, we had placed a mirror at a 45^o angle from the viewing direction, and then photographed the view in the mirror. This reflection reverses left and right. We can say that any point on the left of the symmetry axis of the cathedral has been mapped into an equivalent point on the right, and vice-versa. Physically the cathedral is symmetrical under this transformation (but the image shadows, and the details of the stained glass windows are not).

Reflection symmetry can be found throughout nature. The human body, indeed the human brain, to a good approximation, are bilaterally symmetric. The left-brain and right-brain typically function differently as behavior evolves in an organism, but shape-wise and structure-wise (morphologically) they are the same. Many molecules come in left (levo-) and right (dextro-) forms that differ only by reflection. A mathematical statement of the operation of reflection in a mirror is simply to set up a coordinate system (x,y,z), and describe all of the objects by their coordinates. To reflect the system in a mirror that lies in the xy plane and which passes through the origin, simply replace every z coordinate for every object in the system by its negative, -z. We say that z -> -z is a "reflection in the xy plane."

Rotations:

A sphere (or a spherical system) can be rotated about any axis that passes through the center of the sphere. A cylinder may be rotated through any angle, but only about the special axis of symmetry of the cylinder. A cylinder has the same symmetry as a circle. We can rotate a circle about its diameter, and this is a symmetry operation. The rotation angles we choose can be anything we want, for example, 63^o. After this rotation (often called an "*operation*" or "*transformation*") the appearance of the sphere or circle is not changed. We say that the sphere or circle is "*invariant*" under the "*transformation*" of rotating it about the axis by 63^{o}. Any mathematical description we use of the sphere will also be unchanged (invariant) under this rotation.

Clearly there are an infinite number of symmetry operations that we can perform upon the circle, or the sphere. Furthermore, there is no "smallest" nonzero rotation that we can perform; we can perform "*infinitesimal*" rotations of the circle or on the sphere. We say that the symmetry of the circle or sphere is "*continuous*."

From the many diverse ways of describing symmetry, one eventually gets to an appreciative agreement with the scientists' definition:

"Symmetry is an invariance of an object or system to a set of changes (transformations)."

We have thus encountered two distinct kinds of symmetries,

I. Discrete Symmetries (e.g. Reflections)

II. Continuous Symmetries (e.g. Rotations)

From a mathematician's point of view these are remarkably different kinds of symmetries.

Discrete symmetries are characterized by smallest units of operations, while continuous symmetries have no smallest operation, e.g., we can rotate the circle or sphere through an arbitrarily small rotation angle, or an *infinitesimal* angle. As such, we can use the powerful techniques of differential calculus to analyze continuous symmetries. This leads to an entire branch of mathematics (known as Lie Groups, after the famous mathematician, Sophus Lie who pioneered this technique). All possible continuous symmetries have been classified by the mathematicians. Discrete symmetries have unit "steps" which canot be subdivided; a half of a reflection operation is not a symmetry operation. The complete classification of discrete symmetries was a formidable problem, and has only recently been completed, using computers to prove enormously complicated theorems (See VII. Mathematics of Symmetry).

Let us begin by considering space and time. Space and time contain symmetries, almost obvious yet subtle and even mysterious. Space and time form the stage upon which the dynamics of complicated physical systems is played out. Yet, in a profound way, the symmetries of space and time control the dynamics. This interrelationship between symmetry on the one hand, and dynamics on the other, is the content of the most important logical connection or "*theorem*" we have about nature and the laws of physics: it is called "*Noether's Theorem.*" It is named for its discoverer, Emmy Noether, one of the greatest mathematicians (and, by the way, theoretical physicists) of the 20th century.

What are the Symmetries of Space and Time?

Emmy Noether

Symmetry in Physics: Symmetries of Space and Time

Translations in Space

A physical system can simply be moved from one place to another place in space. This is called a "spatial translation".

Consider a classroom pointer. Usually it is a wooden stick of a fixed length, about 1 meter. We can translate the pointer freely in space. Do its physical properties change as we perform this translation? Clearly they do not. The physical material, the atoms, the arrangement of atoms into molecules, into the fibrous material that is wood, etc., do not vary in any obvious way when we translate the pointer. This is a symmetry: it is a statement that the laws of physics themselves are symmetrical under translations of the system in space. Any *equation* we write describing the quarks, leptons, atoms, molecules, stresses and bulk moduli, electrical resistance, etc., of our pointer must *itself* be invariant under translation in space.

The pointer can be described by giving the locations of its endpoints in space in some coordinate system. We emphasize that the coordinate system is something that we humans cook up to describe things; nature does not come with any coordinate systems of its own. So, let us set up a 3-dimensional orthogonal coordinate system, called A, with an x-axis, a y-axis and a z-axis.

Now we can say that the end of the handle of our pointer is located at the coordinates: (x_1, y_1, z_1). And, the tip of the pointer is located at (x_2, y_2, z_2). A second coordinate system, called B, can be constructed which is translated in space relative to A. The axes of B are parallel to the axes of A, but the origin of B is located at the point (a,b,c) in coordinate system A. Therefore, in B the handle of our pointer is located at (x'_1, y'_1, z'_1) and the tip at (x'_2, y'_2, z'_2). and we have:

$x'_i = x_i + a$
$y'_i = y_i + b$
$z'_i = z_i + c$

Now consider the "physics " of the pointer. The pointer has a physical parameter, which is its "length," R. Can we write an equation for the length of the pointer which gives the same answer in coordinate system A as in system B? In fact, it isn't hard. We say that in any coordinate system the formula for R is:

$$R^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$$

This formula is invariant, or symmetrical, under the translation in space, because we can immediately see that:

$$R^2 = (x_2 + a - x_1 - a)^2 + (y_2 + b - y_1 - b)^2 + (z_2 + c - z_1 - c)^2 = (x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2$$

After performing the translation of the coordinate system we get the same identical formula for the length of the pointer, where the original independent variables (x_i, y_i, z_i) have now been replaced by (x'_i, y'_i, z'_i). Therefore, our formula is valid in all coordinate systems that are translations in space relative to one another. We say that the physical observable, i.e., the *length* of our pointer, is invariant under translations in space.

While this is a simple example, the *highly nontrivial* fact about nature is that *all correct equations in physics are translationally invariant!*

Translations in Time

The physical world is actually a fabric of events. To describe events we need our 3-dimensional spatial coordinate system as described above, but we also need a 1-dimensional coordinate system for time. This is achieved by building a clock. The time t on the clock, together with the position (x,y,z) of something is called an "event" (Note: we always assume that the clock is ideally located at the position of the event, so we don't get confused about how long it takes for light to propagate from the face of a distant clock to the location of the event, etc.).

Some examples of events:

(i) We can say that there was the event of the firecracker explosion at (x_f, y_f, z_f, t_f);
(ii) The N.Y. Yankees' third baseman hits a fast pitch at (x_H, y_H, z_H, t_H);
(iii) Niel Armstrong's foot first touched the surface of the Moon at the event (x_M, y_M, z_M, t_M).

Now, we were actually somewhat sloppy in our discussion of the pointer above, because we treated the ends of the pointer as being located at points in space (x_i, y_i, z_i) but we didn't tell what time it was. I might tell you to meet me at the point (x_m, y_m, z_m), but I must always tell you what time to meet me as well, say t_m. If you show up at (x_m, y_m, z_m) and at any arbitrary time, t'_m, then in all likelihood I will not be at (x_m, y_m, z_m), etc.

When we measure something like the length of the pointer we must specify an event at which that measurement is made.

For example, we can specify that the end of the handle of the pointer is measured at the event (x_1, y_1, z_1, t_1) and the tip of the pointer is measured at the event (x_2, y_2, z_2, t_2) Now, when you measure the length of something, you do this at a common time for the events corresponding to the endpoints of the object you are measuring (Think about it: you have a friend hold down one end of the measuring tape while you position the other end and at the moment you read the marker on the tape the endpoint "events" are simultaneous). So, we need $t_1 = t_2$ and we must slightly enhance our statement about the length of the pointer. We say that the length of the pointer is determined by two formulae together:

$$R^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$$

and:

$$t_1 = t_2$$

The second condition is critical. For example, if the pointer is moving in the +x direction with a velocity v and we measure $x_2 - x_1$ when t_1 is not equal to t_2, then we will actually get an error of $v(t_1 - t_2)$ in our measurement. Hence, we will get the wrong answer for the length of the pointer. *To measure the length we require the endpoint locations to be measured simultaneously* .

Now we see yet another important symmetry of physics: *The laws of physics are invariant under translations in time*. That is, the length of the pointer can be measured now, or tomorrow, or ten seconds ago, etc. If we haven't done anything to the pointer we should get the same answer.

This means that to every time coordinate t_i for every event we are free to add an overall constant representing a time translation, T, and use a new time coordinate $t'_i = t_i + T$. Or, we can say that all of our clocks can be uniformly reset by a constant adjustment of T to the time they read. The laws of physics are invariant under this translation in time. Hence, we see that for the length measurement of our pointer $t_1 = t_2$; therefore, adding T to both sides we also have $t_1 + T = t_2 + T$. Therefore, our definition of the length of the pointer is time translationally invariant:

$$R^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$$

and:

$$t'_1 = t'_2$$

This means that the length of the pointer, as a parameter of an undisturbed (free, inertial) physical system, is invariant under time translations.

The laws of physics, and thus all correct equations in physics, are invariant under translations in both space and time. This is an experimental fact. Indeed, the constancy of the basic parameters of physics over vast distances and times has been established in astronomical and geological observations to approximately 10^{-8} precision.

Rotations

A sphere (or a spherical system) can be rotated about any axis that passes through the center of the sphere. The rotation angle can be anything we want, so let's take it to be 63^o. After this rotation (often called an "operation" or "transformation") the appearance of the sphere is not changed. We say that the sphere is "invariant" under the "transformation" of rotating it about the axis by 63^o. Any mathematical description we use of the sphere will also be unchanged (invariant) under this rotation. There are an infinite number of symmetry operations that we can perform upon the sphere. Furthermore, there is no "smallest" nonzero rotation that we can perform; we can perform "infinitesimal" rotations of the sphere. We say that the symmetry of the sphere is "continuous".

Consider again our classroom pointer. We can rotate the pointer freely in space. Do its physical properties change as we perform this rotation? Clearly they do not. This too is a symmetry: it is a statement that the laws of physics themselves are symmetrical under rotations in space.

Under rotations in free space the length R doesn't change. Suppose we measure the handle of the pointer at the event (0,0,0,0) (using space and time translation we can always choose the handle to be located at the origin of the coordinate system) and we measure the tip end of the pointer at (x,y,z,0) (note that we always choose a simultaneous event for the measurement of length, as described above). Then, in this particular coordinate system the physical length of the pointer is given by our general formula, and becomes:

$R^2 = x^2 + y^2 + z^2$

Now we perform a rotation of the coordinate system which takes (x,y,z,0) -> (x',y',z',0). Note that the rotation doesn't affect the time of any event; we are simply describing the pointer in a new coordinate system that is rotated relative to the old one at the time t=0. The invariance of the pointer's length under this rotation is the statement that:

$R^2 = x'^2 + y'^2 + z'^2$

An example of a rotation is one that lies in the xy plane through an angle of θ:

$x' = x \cos\theta + y \sin\theta$;
$y' = -x \sin\theta + y \cos\theta$;
$z' = z$

(You should convince yourself that the new x' axis, is the line y'= 0 and z'=0; this is rotated by $+\theta$ counterclockwise relative to the old x axis). Now we go ahead and substitute the expressions for x', y' and z' into our equation for R. We easily see that:

$R^2 = x'^2 + y'^2 + z'^2 = (x \cos\theta + y \sin\theta)^2 + (-x \sin\theta + y \cos\theta)^2 + z^2 = x^2 + y^2 + z^2$

(it's not that hard; the $\cos\theta \sin\theta$ terms cancel, and remember that $\cos^2\theta + \sin^2\theta = 1$).

Hence, "observers" using the rotated (or primed) coordinate system write the same equation as the "observers" using the unrotated system for the physical length of the pointer. The length of the pointer is invariant under rotations.

The laws of physics, and thus all correct equations in physics, are invariant under rotations in both space and time . This is, again, an experimental fact.

Emmy Noether

Symmetry in Physics: Special Relativity

IV. Special Relativity

This section can be omitted on a first read of this text. You can skip it and go directly to Emmy Noether's Theorem and Conservation Laws.

Our physical world is a fabric of events, each event labelled by four coordinates, (x,y,z,t). (Remember: Nature does not provide coordinate systems; only humans construct coordinate systems; it's a form of accounting)

In a given coordinate system let us consider two events, $event_1$ at the coordinates, (x_1, y_1, z_1, t_1) and $event_2$ at (x_2, y_2, z_2, t_2). Now, recall from the previous section that rotational invariance is the statement that the *distance* between two points is the same when we rotate our coordiante system. So, the distance between the $point_1$ at (x_1, y_1, z_1) and the $point_2$ at (x_2, y_2, z_2) is

$$R^2 = (x_2-x_1)^2 + (y_2 -y_1)^2 + (z_2-z_1)^2$$

R^2 is the same as in the rotated coordinate system where the points are described by different (rotated) coordinates: $point_1$ becomes (x'_1, y'_1, z'_1) and $point_2$ becomes (x'_2, y'_2, z'_2) and we have

$R^2 = (x'_2-x'_1)^2 + (y'_2 -y'_1)^2 + (z'_2-z'_1)^2$

Rotational symmetry says that R^2 is the same, though the coordinates have changed by the rotation (Remember: the length of the classroom pointer is the same irrespective of its orientation is space) We say the distance is the invariant quantity under rotations. It remains the same before and after the thing we do (rotate), even though the coordinates change under the rotation. Hence, we have a precise statement in mathematical terms of what rotational symmetry means. We also see that our formula for length is invariant under translations, as in the previous section.

Now:

Einstein's Special Theory of Relativity is entirely based upon a symmetry principle that is remarkably similar to that of rotations.

(This is not the usual way that Relativity is introduced to beginning students, so we are taking some risk in explaining it this way. However, it is, in the end, the full content of Special Relativity. And we are using symmetry to try to explain things, so let's try it this novel way.)

The symmetry principle of Relativity states that, a thing very much like the "distance" between the two events, something that is new, and is called the "*proper time*" interval, is the same for all observers. The proper time involves the time separation between two events, as well as the space separation. We will denote by the Greek letter "tau," τ, *the proper time* . It is invariant for all observers, no matter how they are *moving* through space relative to one another (just like the length of the classroom pointer is invariant for all observers no matter how they are oriented in space). That is, τ, *is the same for all inertial observers* (we'll define inertial observers, and talk a lot about the principle of inertia below... bear in mind that we are *never, never* giving up the principle of inertia in Relativity; that is the most important law of physics ever discovered, and it really is the centerpiece of all of Relativity, both Special and General). The definition of the proper time for the two events (1) and (2) described above, the quantity τ, is given by:

$c^2 \tau^2 = c^2(t_2 - t_1)^2 - (x_2-x_1)^2 - (y_2 -y_1)^2 - (z_2-z_1)^2$

Here c is the speed of light (and we haven't even mentioned light! c has more to do with physics than just light! c is a fundamental constant in all of the laws of physics; it essentially calibrates time relative to length in physics, i.e., if we use seconds for time and light-seconds for length measurements, then c = 1.0).

Now, a different observer moving with a constant velocity measures a different set of coordinate values for the two events: event_1 becomes: (x'_1, y'_1, z'_1, t'_1) and event_2 becomes: (x'_2, y'_2, z'_2, t'_2). But, the symmetry principle of Relativity states that τ is the same for all observers:

$c^2 \tau^2 = c^2(t'_2 - t'_1)^2 - (x'_2-x'_1)^2 - (y'_2 -y'_1)^2 - (z'_2-z'_1)^2$

This formula produces the same value of τ in the primed coordinate system as in the unprimed coordinate system. This is the defining symmetry principle of Special Relativity (It's actually called *Lorentz Invariance*). Let us now note some very important implications of our new symmetry, the symmetry of Lorentz Invariance.

Lorentz Invariance Contains Rotational Invariance:

Note that the formula for the proper time involves two pieces, a time piece: $c^2(t_2 - t_1)^2$, and a space piece: $-(x_2-x_1)^2 -(y_2 -y_1)^2 -(z_2-z_1)^2$. We see that the space piece is just the (negative of) the squared distance between the space coordinates of our event, i.e. $-R^2$. Therefore, we can write $c^2 \tau^2 = c^2(t_2 - t_1)^2 - R^2$. So, if we just do a space rotation of our coordinates, τ stays the same. Since Lorentz Invariance is the full symmetry that makes τ the same for all observers, it follows, therefore, that rotations must be a part of Lorentz Invariance, the full symmetry of Relativity!

The proper time is the actual time elapsed on an inertial clock relative to which the events occur at the same place in space:

If the two events occur at the same point in space, then $(x_1, y_1, z_1) = (x_2, y_2, z_2)$. Then the distance between the events becomes zero, or R^2 becomes zero. The proper time then becomes: $c^2 \tau^2 = c^2(t_2 - t_1)^2$, or $\tau = |t_2 - t_1|$. Now, the keyword above is *"inertial"* . We have to specify that our clock is moving inertially, or freely moving through space in the absence of forces, and is not accelerating. Then the time lapsed on the clock is the same as the coordinate time difference. One way to insure this is to say that the two events are occuring so close together in time that the clock can't be dragged around too much in the interim.

However, the "clock" may be attached to an electron (e.g., we may be measuring the electron spin, as if the electron had a little marker on it and we could try to count how many times it spins around per second bound in an orbit around a nucleus, and it may be difficult in practice to keep the motion inertial; this, by the way, is related to something wierd called Thomas precession of electrons in atoms). We'll get to the funny business that Relativity causes for time measurements elsewhere (or go to a good textbook until we update the website to include more about relativity, or wait until we finish our forthcoming book on the subject). Here we are just trying to motivate why we call it the "proper time."

If Lorentz Invariance is the true symmetry of space and time and all of the laws of physics, then Newton and Galileo were wrong!

Now we are finally getting somewhere. This is the big fuss that Einstein created in 1905 when he published his paper "On the Electrodynamics of Moving Bodies," the paper we now think of as the discovery of Special Relativity. What goes wrong with Newton's and Galileo's ideas about inertial obervers?

Let us consider a "stationary" observer (using unprimed coordinates) and a "moving" observer (using primed coordinates). The moving observer has a velocity vector, $V = (v_x, v_y, v_z)$. For the sake of convenience let's specialize to the case of motion in the +x direction, where $v_x = v$, and $v_y = v_z = 0$. Now, according to Galileo and Newton, the coordinates of the event_i for the moving observer are related to the coordinates of the event_i for the stationary observer by the following formulae:

$x'_i = x_i - vt_i$
$y'_i = y_i$
$z'_i = y_i$
$t'_i = t_i$

These formulae are called a *Galilean transformation* and they connect the coordinates used by two different inertial observers. The are much like the connection between observers (primed) in a coordinate system that is rotated relative to another one (unprimed). Notice, they are simple linear formulae. If the moving observer had a velocity v in the +y direction, then we would have $y' = y - vt$. So, the affected space coordinate depends upon the direction of motion. However, notice that time is unaffected by motion in any direction! According to Galileo and Newton, time is *universal* , i.e., it is the same for all inertial observers according to classical physics.

We can now see, however, that

the Galilean transformation is incompatible with the symmetry that requires proper time to be the same for all observers! That is, Galilean transformations violate Lorentz Invariance!

We can easily see this by simply substituting these expressions into the formula for proper time:

$$c^2 \tau^2 = c^2(t'_2 - t'_1)^2 - (x'_2-x'_1)^2 - (y'_2 -y'_1)^2 - (z'_2-z'_1)^2$$
$$= c^2(t_2 - t_1)^2 - (x_2-x_1 - vt_2 + vt_1)^2 - (y_2 -y_1)^2 - (z_2-z_1)^2$$
$$= c^2(t_2 - t_1)^2 - (x_2-x_1)^2 +2v(t_2 - t_1)(x_2-x_1) - v^2(t_2 - t_1)^2 - (y_2 -y_1)^2 - (z_2-z_1)^2$$
$$= c^2 \tau^2 +2v(t_2 - t_1)(x_2-x_1) - v^2(t_2 - t_1)^2$$

And so, we now see that under the Galilean transformation the proper time formula has changed in a peculiar way. The only way Lorentz Invariance could hold would be if

$$0 = 2v(t_2 - t_1)(x_2-x_1) - v^2(t_2 - t_1)^2$$

however, this is not generally the case for arbitrary pairs of events. What we see is that the (squared) proper time changes under the Galilean transformation by an amount:

$$2(v/c)(t_2 - t_1)(x_2-x_1) - (v/c)^2(t_2 - t_1)^2$$

This implies that, for any motion using the Galilean transformation relating the coordinate systems of inertial observers, the symmetry of Lorentz Invariance does not hold. Maybe we can just ignore the discrepancy and say that Lorentz Invariance is only approximately true? Indeed, from a practical standpoint, in the era of Galileo and Newton, an era of horse-back and horse-carriage travel, the discrepancies are very small, of order the ratio (v/c). A typical fast horse travels 10 m/sec, while the speed of light is 3×10^8 m/sec, so this effect is only of order 3×10^{-8}. No experimentalist (even Galileo himself) could have possibly measured such a small effect in those days. For all practical purposes the speed of light was infinite in the era of Galileo and Newton. And indeed, the relevant symmetry of the Galilean transformation is that time intervals are the same for all observers, i.e. $t'_2 - t'_1 = t_2 - t_1$, and , of course, we have rotational and translational invariance. However, proper time invariance is not a symmetry of the Galilean transformation, and therefore, if Lorentz Invariance is to be exactly true then the Galilean transformation must be wrong.

If the Galilean transformation is wrong, then what is the correct transformation beween inertial observers?

We saw above that the simple Galilean transformation is not the symmetry operation that preserves the proper time to be the same for all observers, i.e., the transformation of the symmetry of Lorentz Invariance. Then what is? Clearly we need something that mixes both space and time to cancel the offending terms. Moreover, we would pray to preserve linearity, because a nonlinear formula would cause us severe indigestion. Indeed, the actual symmetry transformation that relates the two inertial observers is just the solution to the problem: "Find the most general linear transformation connecting (x_i, y_i, z_i, t_i) to (x'_i, y'_i, z'_i, t'_i)?" We have already seen that spatial and temporal translations, and rotational transformations will work. However, now there is an additional transformation that relates relatively moving observers, called the "Lorentz Transformation" (Heinrich Lorentz already sort-of had some of these ideas before Einstein, but it was Einstein who first fully appreciated this as a fundamental symmetry principle intertwined with the principle of inertia; in fact, Einstein largely invented our modern way of thinking about symmetry in physics).

The Lorentz transformation corresponding to motion of the primed observers relative to the unprimed ones in the x direction is:

$x' = \gamma\,(x - vt)$
$t' = \gamma\,(t-(v/c^2)x)$
$y' = y$
$z' = z$

where:

$\gamma = 1/(1-v^2/c^2)^{1/2}$

To simplify the algebra, choose $event_1$ at the coordinates, (0, 0, 0, 0) and $event_2$ at (x, y, z, t). (We are always free to define one of the events to be at the origin of our coordinate system because of time and space translational symmetry.) Then the proper time between the events is

$$c^2\ \tau^2 = c^2\ t^2 - x^2 - y^2 - z^2$$

This formula must also hold in the primed coordinates. Thus, we substitute the above expressions for the primed coordinates:

$$c^2\ \tau^2 = c^2\ t'^2 - x'^2 - y'^2 - z'^2$$

and we can using the Lorentz transformation, with a little algebra (go ahead and do it yourself; its easy) verify that we recover:

$$= c^2\ t^2 - x^2 - y^2 - z^2 = c^2\ \tau^2$$

Hence the *Lorentz transformation, indeed, preserves the invariance of the proper time*. Lorentz transformations are correct. They replace the Galilean transformation as the correct symmetry transformation connecting inertial observers.

Note the mathematical similarity of the rotational invariance of the length of the pointer to the Lorentz Invariance of the proper time. The major difference is that the formula for the proper time has + signs in front of terms involving time, and - signs in front of the terms involving distance. This is like the length of the hypotenuse of a right triangle ala the Pythagorean formula, but for the funny sign differences between space and time.

Why is the principle of Lorentz invariance true?

Well, that we cannot answer. We don't know why anything really is the way it is, but we can enlarge our understanding of what it means.

We see that if an event at (0,0,0,0) describes the emission of light, say from an exploding firecracker, then any other event satisfying:

$$0 = c^2 t^2 - x^2 - y^2 - z^2$$

will correspond to the position of the light front (the photons) coming from the firecracker explosion at time $t > 0$. We say that this is the equation of the light-front of photons coming from the source at the event (0,0,0,0). We see that the proper time interval between events connected by light signals is always zero! We call events connected by light signals, "light-like" events. The proper time interval between light-like events is zero.

Einstein's Special Theory of Relativity predicts the fact that all observers measure the same result for the speed of light, no matter how they are moving.

Recall that Michelson and Moreley in 1887 had used the motion of the Earth to try to detect the change in the speed of light when one chases after it. The prediction of the Galilean transformations is that, if you chase after a photon with your velocity being v in the +x direction, where a stationbary observer sees the photon to be traveling at the speed c in the +x direction, then you should observe (c-v) in the +x direction. In principle, you should be able to travel at c and make the photon velocity 0! In principle you should be able to overtake the photon entirely! The Michelson-Moreley experiment produced the result for the speed of light to be c, no matter how much, or in what direction one chased after it. There were many whacky theories around to try to explain this. Someone should write a good book about the history of whacky theories; the later 1890's might be a good place to start.

By starting with the symmetry principle as the definition of Relativity, then the constancy of the speed of light for all observers is predicted. Usually one starts with the assumption that the speed of light is the same for all observers and derives the Lorentz transformation. Einstein started with the assumption that Maxwell's equations (the equations that describe electromagnetism, and therefore light) are the same for all observers and that the parameter c, which enters Maxwell's equations, is the same value for all observers. In a sense, this is the only natural way to think about Maxwell's equations, so Special Relativity is encoded into them. However, the truly fundamental point underlying all of this is the symmetry of Lorentz Invariance: the fact that proper time between any pair of events is the same for all observers. That is the true essence of it all.

Hence, all observers must write down, in their respective coordinate systems, the same equation for the light-front, with the same universal velocity of light c. So, our moving observer must write, in her coordinate system:

$$0 = c^2 t'^2 - x'^2 - y'^2 - z'^2$$

The fact that τ is the same for all observers, under the Lorentz transformation that relates them, guarantees that this will also be a correct equation of motion for any moving inertial observer. All observers will agree on one universal constant speed of light. You will always observe the universal speed of light, c, for a photon no matter how you chase after it. Remarkably, Einstein replaced the universality of time by the universality of the speed of light! Time measurements are therefore observer dependent.

We see that as the velocity of the moving observer approaches the speed of light, i.e., when v -> c, the factor γ blows up and the formulae of the Lorentz transformation break down. This is symptomatic of a general property of Relativity: things just can't go faster than the speed of light!!! We will see later how energy and momentum also become infinite in this limit.

Many people have speculated about the possibility of objects, called tachyons, that hypothetically do travel faster than the speed of light. There is no convincing theory of these objects, and certainly no experimental evidence for them. In fact, tachyons are usually associated with an unstable vacuum in quantum theories, and though certain instabilities can produce apparent super-luminal things (like tidal waves of vacuum rearranging fields) no physical signals ever end up going faster than the speed of light. Moreover, all of this is tightly interwoven in the quantum theory to produce a remarkable new phenomenon called "anti-matter." Antimatter was actually predicted before it was discovered by Paul Dirac, who combined Relativity and Quantum Mechanics into one theory (called Relativistic Quantum Mechanics). Later, after Noether's theorem, we will explain this to you, as well. Take our word for it, anti-matter exists, it is real, and we make it and use it in physics laboratories like Fermilab every day (at least when we have the $$$ to run the machine, which isn't everyday these days). At present we don't know what good anti-matter is for the general economy, but we do know that one day the government will tax it.

We also see that the "proper time" in the limit c -> infinity, becomes just the time interval between events without any reference to space separation: τ -> t. So, as c -> infinity, time becomes absolute and disconnected from space; all observers can use the same time coordinate for all events in this limit. Note, too, that if you take c -> infinity in the formulae for the Lorentz transformation we get back to the Galilean transformation:

x' = x-vt
t'=t, etc.

Hence, there was Relativity in Galileo's day though few people back then really understood this in terms of symmetry. It should be noted that the Principle of Relativity is the "equivalence of all inertial frames (uniformly moving coordinate systems) for the formulation of laws of physics." Hence it really underlies the law of inertia.

All of classical physics (pre-relativity, pre-quantum) is based upon Galilean Invariance. Relativity subsumes this into Lorentz Invariance. All of this begins with the idea of *inertial observers* . Inertial observers are non-accelerating, freely moving without any external forces acting upon them. Both the symmetry of Galilean and of Lorentz Invariance states that any two inertial observers will find that the laws of physics are the same for them, no matter how fast (or in what direction) they are moving relative to one another.

Now this principle underlies in a deep way the principle of inertia. The principle of inertial states:

A body at rest will remain at rest *unless acted upon by an external force*.

A body in uniform translational motion will remain in that state of motion *unless acted upon by an external force*.

Stated this way it appears as though there are two special states of motion, i.e., " *rest* " or no motion, and *"uniform translational"* motion (straight line motion with no acceleration, or bumps or jars). However, Galileo, Newton (and this principle is often credited earlier to Rene Descartes as well) understood that there really is no difference between a state of rest and a state of uniform translational motion. That is, it is really a matter of the observer's state of motion as well. If you observe an object at rest a and if I am moving at a velocity v in the +x direction, then I will observe the object as moving with velocity -v in the +x direction relative to me! If physics is the same for you and me, then we both conclude the object is either at rest or in a state of uniform motion if not acted upon by a force! Bottom line: There is no absolute state of motion or rest.

The principle of inertia therefore implies that all observers observe the same laws of physics to hold true independently of their state of relative motion, at least in the absence of forces However, if forces are produced by other physical systems, then forces too must be governed by this principle. So, all inertial observers must agree that the laws of physics are the same for them, and motion must be a symmetry operation like rotation that leaves the laws unchanged. (Note: this does not mean that the observers will observe the exact same thing; time, length, energy, momenta, etc, for a given physical event will appear to be different for different observers; It is the laws of physics that interrelate these quantities that will always appear to be the same!)

We almost take these basic notions of inertia, which begin in Newtonian and Galilean physics, and which apply to this day in all of physics, for granted, or as self-evident. However, they are far from self-evident. The Greeks never conceived of the idea that an object in motion remains in motion unless a force acts upon it. The Greek's, and indeed, all of the natural philosophers (physicists) down to the era of Galileo had it all wrong. They all thought that *an object moved only if an external force was applied* . Take away the force, and the object has a "natural tendency" to come to rest. Even the great Johannes Kepler, who gave a set of laws of motion for the planets that encoded the correct results of the Newtonian theory of gravity to come later, thought that something was pushing the planets about in their orbits. In some sense, the relationship between force and motion that the Greek philosophers had in mind was:

"Force = (Mass) x (velocity)"

Obviously, this view causes one to completely misunderstand the nature of the forces at work that produce planetary motion. If the force vector is tangential to the orbit then it is hard to see what the sun has to with it. However, once one realizes the principle of inertia (which may have originated with Rene Descartes) that the velocity is constant unless a force is applied, then the next best guess is Newton's:

"Force = (Mass) x (acceleration)"

For a planet moving in a circular orbit, the acceleration is directed toward the center (it is "centripedal"). Now the correct formula ala Newton allows us to see that the force must also be directed toward the center. Voila, a possible connection to the sun as the source of the force now becomes possible!

Unfortunately, we must now send you off to a textbook on Special Relativity to learn about all of the miraculous effects that occur as a consequence of the Lorentz transformation. We will over time develop a more complete treatment of Relativity for the website. Indeed, Special Relativity is conceptual, but involves no more mathematics than algebra (it doesn't involve Calculus until it is applied to more complex situations).

Our key point here is that Relativity is an expansion of our understanding of the deep secrets of the symmetry of nature. We'll see below, however, how it all leads to the existence of anti-matter!

The laws of physics are invariant under spatial and temporal translations, rotations in space, and Lorentz transformations of space-time.

Emmy Noether

Symmetries of the Laws of Physics and Noether's Theorem

In 1905, a mathematician named Emmy (Amalie) Noether* proved the following theorem:

For every continuous symmetry of the laws of physics, there must exist a conservation law.

For every conservation law, there must exist a continuous symmetry.

Thus, we have a deep, deep connection between a symmetry of the laws of physics, and the existence of a corresponding conservation law. In presenting Noether's theorem at this level we usually state it without proof. A simple proof can be given if the student is familiar with the action principle. However, it is better to motivate the result through examples. (For a proof, check out Mathematics of Symmetry and

Physics.)

Conservation laws, like the conservation of energy, momentum and angular momentum (these are the most famous), are studied in high school. We now see from Noether's theorem that they emerge from symmetry concepts far deeper than Newton's laws. We will also learn that there are many other conservation laws in physics. Finally, we will give some idea of how this theorem plays out in the quantum theory domain.

Now, as we have stated above, it is an experimental fact about the nature that the laws of physics are invariant under spatial translations. This is a strong statement. For example, if space had the structure of a crystal, then moving the origin of coordinates from a nucleus to a void would change the laws of nature within the crystal. The hypothesis that space is translationally invariant is equivalent to the statement that one point in space is equivalent to any other point, i.e. the symmetry is such that translations of any system or, equivalently, the translation of the coordinate system, does not change the laws of nature. Equivalently, the laws (and equations that express these laws) are invariant to translations (translational symmetry).

Now comes the amazing result of mathematician Emmy Noether, whose theorem, in this case, states:

The conservation law corresponding to space translational symmetry is the Law of Conservation of Momentum.

So, we learn in senior physics class that the total momentum of an isolated system remains constant. The *i*th element of the system has a momentum in Newtonian physics of the form:

$$p_i = m\ v_i$$

and the total momentum is just the sum of all of the elements,

$$P_{\text{total}} = p_1 + p_2 + \ldots + p_N$$

for a system of N elements. Noether's theorem states that P is conserved, i.e., it does not change in time, no matter how the various particles interact, because the interactions are determined by laws that don't depend upon where the whole system is located in space!

Note that momentum is, and must be, a vector quantity (hence the little arrow over the stuff in the equations). Why? Because momentum is associated with translations in space, and the directions you can translate (move) a physical system form a vector! So, if you remember the Noether theorem, you won't forget that momentum is a vector when taking an SAT test!

Turning it around, the validity of the Law of Conservation of Momentum as an observational fact, via Noether's theorem, supports the hypothesis that space is homogeneous, i.e., possessing translational symmetry. The more we verify the law of conservation of momentum, and it has been tested literally trillions of times in laboratories all over the world, at all distance scales, the more we verify the idea that space is homogeneous!

We note that Noether's Theorem further assures us that for any translationally invariant physical system there is always something called "momentum" and it is always conserved. The exact formula for the momentum depends upon what we are studying. For the Newtonian particle it takes the form mv, while for the relativistic particle, $mv/(1- v^2/c^2)^{1/2}$, and for the electromagnetic wave $E \times B/4\pi c$, etc. Note that in each case the momentum is always a vector (3 component) object corresponding to the three translational directions of space.

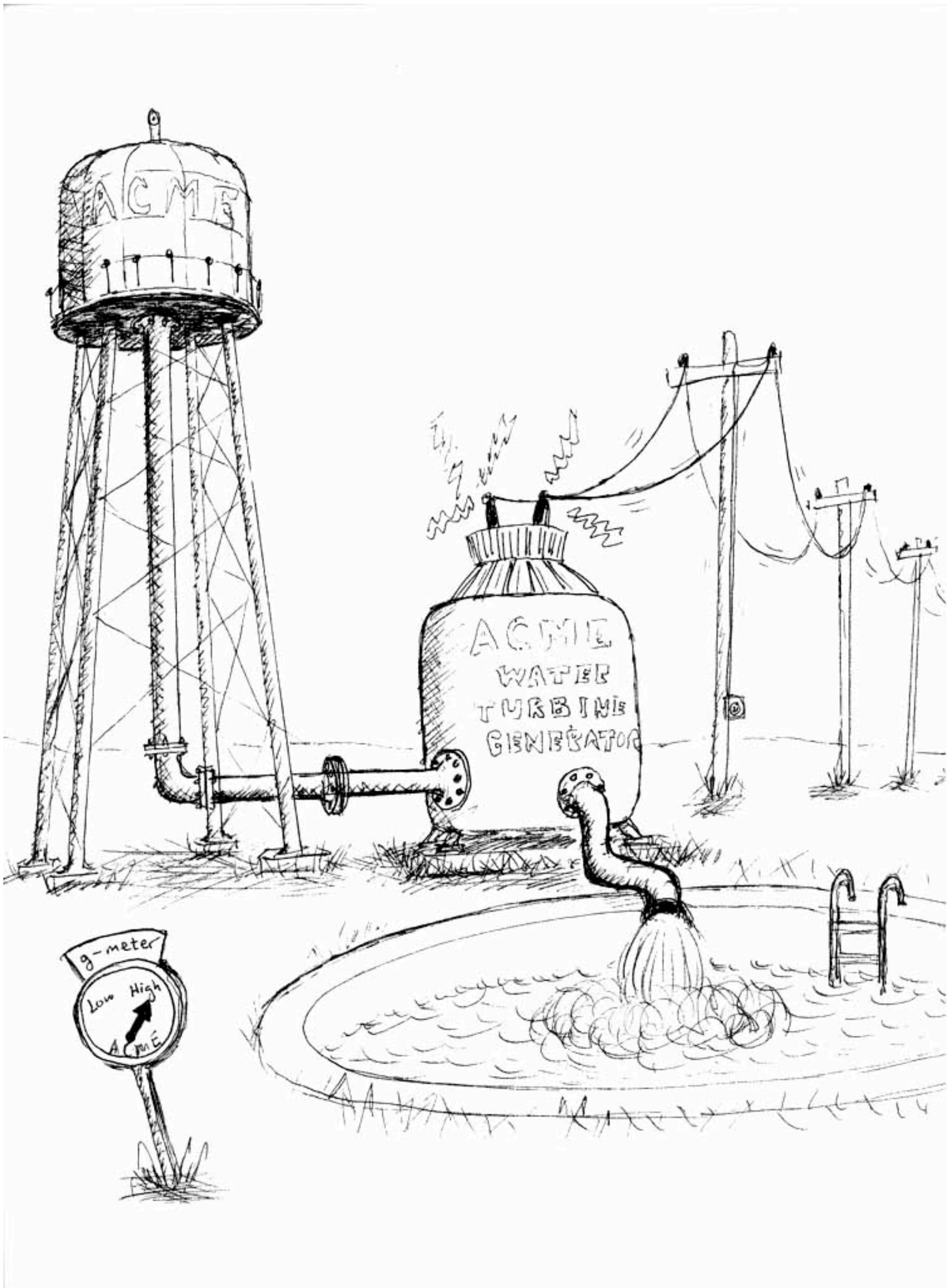
The experimental evidence also favors very strongly the homogeneity of time, i.e. any point on the time axis is as good as any other point, i.e., the laws of physics are invariant under tanslations in time. What conservation law then follows by Noether's Theorem?

Surprise! It is nothing less than the law of conservation of energy:

The conservation law corresponding to time translational symmetry is the Law of Conservation of Energy.

Since the constancy of the total energy of a system is extremely well tested experimentally, this tells us that nature's laws are invariant under time translations.

Here is an example of how time invariance and energy conservation are interrelated. Consider a water tower that can hold a mass M of water and has a height of H meters. Assume that the gravitational constant, which determines the acceleration of gravity, is g, on every day of the week except Tuesday, when it is a smaller value $g' < g$. Now, we run water down from the water tower on Monday through a turbine (a fancy water wheel) generator which converts the potential energy MgH to electrical current to charge a large storage battery. We'll assume 100% efficiencies for everything, because we are physicists. This is Monday's job.

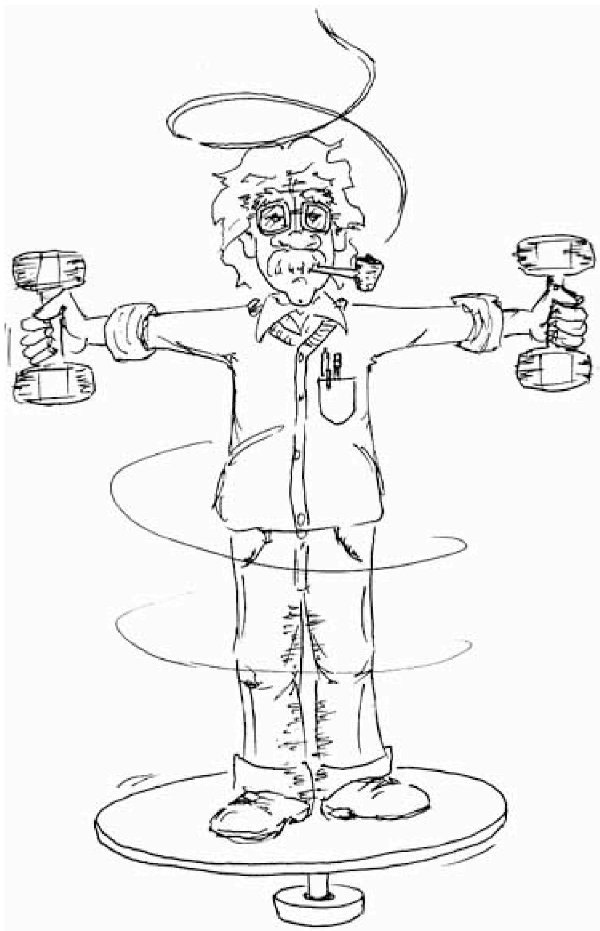


Now on Tuesday we pump the water back up to H, using the battery power that we accumulated from Monday's job to run the pump. But now the g' value is smaller than g and the work done is Mg'H, which is now much less than the energy we got from Monday's job. This leaves us with M(g-g')H extra energy still in the battery, which we can sell to a local power company to live on until next Monday. This is a perpetual motion machine! It produces energy for us, and we can convert that to cash. It does not conserve energy because we cooked up false laws of physics, in this case gravity, that are not time translationally invariant!!! Hence, we violated a precept of Noether's Theorem. (Can you come up with similar cute example of violating momentum conservation by making the laws of physics spatially inhomogeneous?)

We also live in a world where the laws of physics are rotationally invariant:

The conservation law corresponding to rotational symmetry is the Law of Conservation of Angular Momentum.

Conservation of angular momentum is often demonstrated in lecture by what is usually called "the 3 dumbbell experiment". The instructor stands on a rotating table,



his hands outstretched, with a heavy dumbbell in each hand (who is the third dumbbell?). A student starts the lecturer rotating on his table. He turns slowly, then brings his hands (and dumbbells) close to his body. His rotation speed (angular velocity, ω) increases substantially. What is kept constant is the angular momentum, J, the product of I, the moment of inertia, times the angular velocity \omega. Hence, J = Iω

By bringing his dumbbells in close to his body, the moment of inertia I, is decreased (roughly speaking, the moment of inertia is I_0, the Professor's body's moment, plus the two dumbell masses 2M times the length of his arms squared, r^2, or $I = I_0 + 2Mr^2$; in the experiment, r is initially the extended length of his arms, and the finally is approximately 0, the retracted length of the arms). But J, the angular momentum, must be conserved, so ω must increase. Skaters do this trick all the time.

Atoms, elementary particles, etc., all have angular momentum and in any reaction, the final angular momentum must be equal to the initial angular momentum. Like our planet earth, particles spin and execute orbits and both motions have associated angular momentum. Data over the past 70 or so years confirms conservation of this quantity on the macroscopic scale of people and their machines and on the microscopic scale of particles. And now, (thanks to Emmy) we learn that these data imply that space is isotropic; —there is no preferred direction. All directions in space are equivalent. Incidently, the conservation of angular momentum is encoded into Kepler's third law of planetary motion, and in some sense represents the first exact statement of a conservation law (Archimedes anticipated energy conservation).

These translational and rotational symmetries of space and time need not have existed. That they do is the way nature is. We are learning some of the actual properties of the concepts we use to describe the world: space and time. It didn't have to be this way. For example, if all of space were constructed like the insides of a crystal, then all directions would not be equivalent, and continuous translational symmetry would be lost.

What about the Lorentz invariance? What is the conserved quantity associate with it? Actually, the devil gets into this one; we find that the conserved quantity is actually 0. Fortunately, 0 is conserved, so there is a technical conservation law here, it just isn't a useful one. Nonetheless, Lorentz invariance has profound effects, like the fact that mass is equivalent to energy through $E=mc^2$, which we'll prove below, using Noether's Theorem.

There are more abstract symmetries which do not involve space or time coordinates. An example is the symmetry of an assemblage of positive and negative charges. We can define an operation that tests for symmetry as follows: "change the signs of all the charges." In this case the appearance of the system changes. For example, if we have a Hydrogen atom, the nucleus is a positively charged proton and the electron orbiting the nucleus is negatively charged. After performing our operation, we are left with a system containing a negatively charged nucleus and a positively charged electron orbiting. The "invariance" under this operation is the statement that both systems have identical physical properties.

Such a system could in principle be constructed. Positive electrons were discovered in 1932 (positrons) and negative protons were discovered in 1955. These are examples of "antimatter." An anti-Hydrogen atom could be constructed from the anti-proton and the positron; in fact, it has been made in a clever particle accelerator experiment, and some of its properties have been observed. Matter-anti-matter symmetry implies that the mass, spins, binding energies, excited states, etc. of both real Hydrogen and anti-Hydrogen must be identical. The charge-reversal symmetry is another example of a discrete symmetry. We will have more to say about anti-matter, and ultimately we can glimpse why anti-matter must exist from the basic symmetry principles of space and time embodied in Einstein's Theory of Relativity.

*(pronounced like "mother" with an "n"; born, 1882, she had a great deal of trouble finding permanent employment in male-dominated European universities, and had to flee the rise of Naziism; she spent her last few years at Bryn Mawr; she died in 1935)

Emmy Noether

Symmetry in Physics: Symmetries in Quantum Mechanics

We would like to illustrate the generality and power of some symmetries that emerge in the quantum theory uniquely. These are the symmetries of "phase invariance" or "local gauge invariance," and the symmetries associated with interchanging "identical particles". These symmetries play a profound role in controlling the structure of the basic forces in nature, the conservation of particle number, the conservation of electric and other charges, etc. etc. These symmetries ultimately control nature as we see it and feel it, from controlling the basic stability of matter, dictating the Periodic Table of the Elements, to producing wierd phenomena such as lasers, superconductors, superfluids, etc. etc.

First we must approach the question: what is quantum theory?

The Wave Function

Suppose we have only one particle in the whole Universe. Of course, this is always just an approximation in which one particle is treated approximately in isolation from everything else. In fact, it is a pretty good approximation for freely propagating light (photons) or electrons or neutrons or protons or atoms (viewed as particles) straying freely in space (and even to a lesser degree within a metal) and even for quarks found inside the proton at extremely short distances (or in very very high energy collisions).

How do we describe such a particle? In quantum mechanics we do so by introducing the idea of the "quantum state." A particular form of this idea is to talk about the "wave-function" of the particle. And, for simplicity, we will assume that our particle is extremely elementary and carries no information other than where it is located. [in a vague sense that is what a particle is; it is a tiny object which carries a limited amount of information, such as (1) its position, (2) its "spin" i.e. how much angular momentum resides with the particle, and this is subject to special rules from quantum mechanics, (3) its electric charge, and various other tidbits of information that we won't consider at present; however, if you are a computer geek you can think of an elementary particle as a kind of "byte" of information]. For our very-elementary particle that carries only information about position the wave-function is simply a (complex number valued) function of space, and time:

$\psi(x, t)$

The meaning of the wave-function is subtle and profound, and involves the full construction of the quantum theory. We will tell you however, that it is useful, because the (absolute) square of the wave-function is the probability of finding our particle at any given point in space, at any particular time:

$|\psi(x, t)|^2$ = probability of finding particle at position x at the time t.

Let us emphasize that the wave-function for particles like electrons associates a complex number (a number of the form (real number)$_1$ + *i*(real number)$_2$, where *i* is the infamous "square root of minus 1"). To form the observable probability of finding the particle somewhere at sometime, note that we take the "absolute square" (the absolute square of a complex number z is z times its complex conjugate z^*, or $|z|^2 = zz^*$.)

At this point you students say: "Surely you jest!!!" "Don't you mean that you are merely using complex numbers for a kind of convenience, like they do in electrical engineering, and that there really is no physical significance to the use of complex numbers?"

To wit we answer: No!!! We do not Jest!!! In Quantum Mechanics there really are complex numbers and the wave-function really is a complex valued function of space-time. Now, *of course* we could reduce everything to ordered pairs of real numbers, and do all the math without ever talking about *i*, but there is no benefit or meaningful significance to doing so. That would be like talking about some horrible social disease at a cocktail party without ever using the actual name of the disease, but everyone would still understand what we were really talking about, and some character might sooner or later blurt it out. The fact is, in the mathematics of quantum mechanics the square-root of -1 plays a fundamental role. Nature reads books on complex numbers!!!

However, since in the end we measure real numbers, i.e., we measure probabilities like $|\psi(x, t)(x)|^2$, we might ask: "what is the significance of the overall complex phase of the wave-function?" Do we smell a symmetry here? What happens if we take the electron wave-function and multiply by an overall complex phase factor?

$\psi(x) \to \exp(i\theta)\,\psi(x)$

Now the probability is unchanged:

$|\psi(x)|^2 \to |\psi(x)|^2$

This is, therefore, a symmetry operation in quantum mechanics.

Now the mind boggling, goose-pimple inducing, amazing consequence of this symmetry is the corresponding conservation law (remember, Emmy Noether says for every continuous symmetry there is a conserved quantity; we have here a continuous symmetry because the real number θ can be any value we like; this is, by the way, the U(1) symmetry we encountered at the very beginning of these lectures). If the particle we are talking about is actually the electron, then we have discovered nothing less than:

The conservation of electric charge!

The conservation of electric charge is a very well tested experimental law of physics. For example, it forbids processes like:

e^- -> neutrino + photon,

in which case electric charge would completely disappear. Indeed, we never see processes like this in the laboratory.

On the other hand, processes like this one do occur:

e^- + p^+ -> n + neutrino

Since the final state is electrically neutral, the negative electric charge of the electron has combined with the positive electric charge of the proton, and by electric charge conservation, we are assured that the electric charge of the electron is equal to the opposite of that of the proton to an infinite number of significant figures. Indeed, we can place a large quantity of Hydrogen gas into a container and obeserve to a very high precision that Hydrogen atoms (which are just bound states of e^- + p^+) are electrically neutral.

Now for some tricky questions:

Q: If photons are also described by wave-functions, then why aren't photons charged?
A: Now here is where the complex numbers are so important. The wave-functions of photons are always real numbers, and therefore we cannot multiply by an overall phase factor, and ergo photons carry no electric charge!

Q: Oh, so things like neutrons also have real wave-functions?
A: NO, neutrons are different; they also have complex wave-functions, and you can multiply by the complex phase factor just like for electrons but it can be a different phase factor. For neutrons there is also a conserved charge, as there must be according to Emmy Noether's theorem, but now it is not electric charge. For neutrons the conserved charge is called "baryon number" and it doesn't involve electromagnetism.

Q: So how can I tell if a given complex wave-function has electric charge or some other charge, like baryon number, associated with it?
A: Smart kid! Well consider the following...

In fact, the symmetry associated with electric charge is even larger than just multiplying the electron wave-function by an overall constant phase. In fact, it is infinitely larger. We can actually multiply the electron wave-function by a phase factor that is an arbitrary function of space-time:

ψ(x, t) -> exp(i θ(x, t)) ψ(x, t)

This gives not only the conservation of electric charge, but as a symmetry of nature we find that we must mathematically include all of the electromagnetic physics, i.e., we need the photon and the precise equations of motion of photons, which leads to the Maxwell equations of electromagnetism , etc. This simple but enormous symmetry, is called "local gauge invariance" and it actually defines all of electrodynamics!!! The only missing parameters we must include are the value of the electric charge and the mass of the electron.

The local gauge invariance distinguishes the symmetry operation we perform on an electron (it is a function of space-time) from that we do to a neutron (it is only a constant). We call the latter case a "global gauge transformation" and it is a very much humbler symmetry than the "local gauge transformation" we do on all charged particle wave-functions. In fact, some people think that true global charges cannot exist at all in physics, and that only local charges can exist. This comes from the idea of itsy-bitsy-mini-tini-black-holes that form part of the vacuum itself at very very short distances. Black-holes can swallow global charge and make it disappear. Hence, baryon number would fall into a black-hole and we would seem to see a loss of net baryon number in a given process. Electric charge, however, cannot be swallowed by a black-hole because when a black-hole swallows an electron, the black-hole itself becomes electrically charges and electric charge is again conserved!!!. So, we think that only local gauge symmetries are true symmetries of nature, while global ones like baryon number, are fakes. So, if a particle has a charge q (electron has q = -1) the rule is:

ψ(x, t) -> exp(-iq θ(x, t)) ψ(x, t) <--> Electromagnetic Gauge Transformation

We should put the above equation on all Tee-shirts and then say "Let There Be Light."

Now, we won't get into it here, but it turns out that all of the forces in nature come from generalizations of this important symmetry:

All forces come from Local Gauge Symmetries

This includes electromagnetism (local phase invariance), gravity (local coordinate system invariance), the strong force (associated with quark color, and the freedom to rotate a quark in color space locally), and the weak interactions (involving rotations between charged leptons and neutrinos, or between up quarks and down quarks).

This has been a brief descriptive expose of some of the headiest issues in modern physics. We have briefly toured the frontier. If you are interested in some more of this, it will be described in the mathematical soliloquys.

Identical Particles in Quantum Theory

Do students who study the (Mendeleev) Periodic Table wonder why the elements go through a cycle of properties? For example, Lithium is an active metal--but as we count off to the right of the table, we get to Neon, which is an inert gas. Why? The general statement is that Lithium has a single electron in a "shell" and Neon has a complete, and hence, a closed shell. But why? Why shells? What rules cover the arrangement of electrons as we proceed from element to element?

We start by assuming that, as we go progressively to more positive nuclei (with Z protons) and add electrons to keep the atom neutral, nature seeks the lowest energy state. We know that the states of atoms are quantized: but that atoms will organize themselves to occupy the lowest energy state. Since one gets less energy by moving electrons close to the nucleus, why aren't all the added electrons as close to the nucleus as possible? The answer is: They are, but exchange symmetry forces electrons to keep away from each other. A new "effective force" appears, which is, in fact, much more powerful than the Coulomb repulsion of like sign electric charges. This "exchange symmetry" plays a crucial role in the formation of solids, atoms, nuclei and hadrons (systems containing quarks), as well as neutron stars, white dwarfs, and black holes . . . but we digress!

Consider a simple physical system containing two electrons, for example, a Helium atom. We describe the Helium atom by a quantum mechanical wave-function that depends upon the positions of two electrons (for the moment we will simplify things and assume that electrons are only described by their positions; this is false because they also have another associated quantity called spin):

$\psi(x_1, x_2, t)$

This is a function of the positions of electron (1), x_1 and electron (2), x_2 and time t.. Again, it is useful, because the (absolute) square of the wave-function is the probability of finding our electrons at the points in space at the given time:

$|\psi(x_1, x_2, t)|^2$ = probability of finding electrons at x_1 and x_2 at time t.

Now consider the act of exchanging one particle with the other particle. In other words, we rearrange our system with the swapping of the positions (and, if we kept the extra information, we would also swap spins, etc.)

$x_1 <--> x_2$

Hence the new "swapped" system is described by the wave-function:

$\psi(x_2, x_1, t)$

But is this really a new system or just the original system we started with?

Now, whereas the class of things called "dogs" is large, no two dogs are identical, even if they happen to both be poodles. However, all electrons are precisely identical to each other. Electrons, remember, carry only a very limited amount of information (and for the moment we are assuming that it is only positional information). Electrons do not have freckles, warts, surgical marks, dental records, or other identifying body markings. Any given electron fresh from the electron factory is exactly identical to any other electron!

Therefore, any physical system must be invariant under the swapping of one electron with another . In a sense, nature is simple minded in the way it treats electrons in that it doesn't know the difference between any two (or more) electrons in the whole Universe. At the quantum level this amounts to a true reduction in the information content of the Universe from that which we would have carried over from Classical Physics! This is also true of replacing one oxygen atom, or one iron nucleus, or one benzene molecule, by another. The strong physics content of this symmetry is just that all protons are identical to one another, as are all electrons, neutrons, etc. Combinations, like molecules H_2O are identical.

This "exchange symmetry" of the wave-function must leave the laws of physics invariant because the particles are identical. At the quantum level it implies that our exchanged wave-function is indistinguishable from the original one:

$\psi(x_2, x_1, t) = (+ \text{ or } -)\psi(x_1, x_2, t)$

Notice that the exchanged wave-function can in principle, equal either + or - times the original one. This is allowed because we can only measure probabilities (squares of wave-functions). So, which is it , + or -?

Now, suppose that instead of electrons we were talking about photons or mesons (particles called "bosons" with no spin, or spin = 1, or spin = 2 , or any integer value of spin). Then the rule is that under swapping two particles in the wave-function we would get the + sign.

$\psi(x_2, x_1, t) = (+)\psi(x_1, x_2, t)$ for BOSONS

We can now prove an important theorem: This means that two identical bosons can share the same point in space, i.e.,

$\psi(x_2 = x, x_1 = x, t)$ need not be zero!

In fact, from just a probabilistic point of view we can prove that if the two particles can share the same point in space, it is possible to coax a lot of them to share the same point in space, and thus bosons like to condense down into compact "coherent" states. This is called Bose-Einstein condensation. It occurs in many places in the fabric of nature, and is one of the many miracles of quantum mechanics.

There are all kinds of variations on Bose-Einstein condensation and all kinds of phenomena that are quite similar:

Lasers produce coherent states of many many photons all moving together in exactly the same state of motion at the same time.

Superconductors involve pairs of electrons bound by crystal vibrations (quantum sound) into spin = 0 states called Cooper pairs. In a superconductor the electric current involve a coherent motion of many many Cooper pairs sharing exactly the same state of motion

Superfluids are states of extreme low temperature bosons (like in liquid He^4) in which the entire liquid condenses into a common state of motion which becomes completely frictionless.

Bose-Einstein condensates have recently been created in the laboratory in which many bosonic atoms condense down into ultra-compact droplets of very large density, and move together in a common quantum mechanical state.

For electrons, on the other hand, we get the (-) sign. We will "explain" this to you in a mathematical soliloquy later on. It has to do with the fact that electrons have spin = 1/2 and are called "fermions" (any particle with fractional spin, 1/2, 3/2, etc, is a fermion).

$\psi(x_2, x_1, t) = (-)\psi(x_1, x_2, t)$ for FERMIONS

Basically a spin = 1/2 particle is described with angular momentum that is a square-root of a vector (this is called a spinor). When we swap the positions of two fermions, it is like rotating the system in space by 180 degrees. Because of the square-root of the vector spins, this rotation produces a net (-) sign when we swap the pairs of electrons.

We can now prove a theorem about all fermions: No two identical fermions (with the same spins i.e., with spins "aligned") can occupy the same point in space:

$\psi(x_2 = x, x_1 = x) = - \psi(x_1 = x, x_2 = x) = 0!$

More generally, no two identical fermions can occupy the same quantum state. This was first known as the "Pauli Exclusion Principle," after the Austrian-Swiss theorist Wolfgang Pauli. Pauli also proved that it comes from the basic symmetries of the laws of physics, though few theorists understand this proof in detail; its much harder than the Noether theorem; we give a sketch of it in the mathematical soliloquy that is due to R.P. Feynman).

This property of fermions, particularly electrons, largely accounts for the stability of matter. For example, in a Helium atom we can get two electrons into the same state if one has its spin pointed "up" and the other with spin "down" (for spinors, the "up" and "down" spins are actually orthogonal(!), i.e., they share no common components). However, we cannot then insert a third electron into that state because its spin would overlap with one of the two electrons already present, and the exchange symmetry minus sign would give zero if we try to construct such a state. Hence, for the atom Lithium, the third electron must go into a new state of motion, i.e., a new orbital. Thus, Lithium has a closed inner orbital or "shell" (a Helium state inside) and a sole outer electron that behaves much like the sole electron in Hydrogen. Lithium and Hydrogen therefore have very similar chemical properties. We thus begin to see the emergence of the Periodic Table of Elements!

The build-up of the elements from Hydrogen to Uranium is dominated by this exclusion principle. As we add more electrons to make higher atoms, they cannot all go to the lowest energy orbital of Hydrogen, because it is already filled with electrons and the new ones we add are excluded. So the electrons must populate increasingly higher energy orbitals until that particular "shell" is filled. Each shell has room for a certain number of electrons. The number of electrons in the outermost shell determines the chemical properties of each atom. All of this is a consequence of exchange symmetry!!!

Yet another extreme example is that of the neutron star. A neutron star is formed as the core of a giant supernova implodes while the rest of the star is blown to smitherenes into outer space. (We were all once residents of the interior of a gigantic star that cooked up our heavy elements, exploded, and allowed the reformation of the present day solar system!) The neutron star is made entirely of gravitationally bound neutrons. Since neutrons are fermions with spin = 1/2, the exclusion principle applies, and the state of the star is supported by the fact that it is impossible to get more than two neutrons (each with spins counteraligned) into the same state of motion. If we try to compress the star the neutrons begin to increase their energies because they cannot condense into a common low energy state. Hence, there is a kind of pressure, or resistance to collapse, driven by the fact that fermions are not allowed into the same quantum state.

Neutron stars are strongly believed to exist. In fact, a neutron star often traps the magnetic field of the supernova that produced it. This intense magnetic field then rotates with the neutron star at a high frequency, maybe hundreds of times per second. This in turn produces the phenomenon of rapid flashes of light seen emanating from the star, known as a "pulsar." Many, many pulsars have been discovered.

Well, that's what science says in any case. An alternative explanation of pulsars is that they are the communication beacons of some large interstellar cell-phone network of an advanced extra-terrestual civilization.

Remarkably, if the mass of the neutron star exceeds about 1.4 solar masses, then gravity can actually win. However, gravity only wins by making one of the most extreme objects anyone has every thought of: a black hole.

Pause: Remember, all of this comes from the exchange symmetry of the quantum wave-functions of elementary particles!

The apparent failure to observe this exchange symmetry in any obvious way in the case of poodles or people or any other everyday macroscopic object is "simply" a consequence of the complexity. Complexity requires that the individual particles have to be far apart from one another, so that many different states are possible, and the particles don't come close to being in the same quantum states at the same time. Thus the effects of exchange degeneracy are not obvious in complex extended systems.

Mirror Symmetry (Parity)

Now for the symmetry demonstrated by the process of reflection in space. Suppose one wall of our ultra laboratory is a mirror. We see two views—the "real" lab, with all kinds of research, measurements, data logs, etc., and a similarly active, similarly precise set of activities in the mirror. If a video were taken of the two views, could anyone tell the difference?

Mirror symmetry, or reflection symmetry (or parity), is the statement that both videos represent the way the world is; i.e. there is no physical (or biological) system or physical process whose mirror image isn't also a possible physical system or process. We would then say that the laws of nature are invariant to reflection in space. In the simplest example of a plane mirror, we replace all distances to the reflection plane, z by $-z$.

Now a box of screws in the lab would usually be right-handed screws, i.e. a clockwise rotation moves the screw forward. Seen in the mirror, the clockwise rotation becomes counter-clockwise, but the mirror image of the screw also moves forward, i.e. the mirror screw is left-handed. So? The point, however, is that a left-handed screw can exist in the lab, it just takes a special order (e.g. "please make us 10 dozen 8-32 left-handed screws"), and so the mirror lab just has different conventions.

Men's jackets have buttons on the right side in the lab, but on the left side in the mirror—again convention. Since mirror reflection changes a right-handed effect into a left-handed effect, it is equivalent to the statement that nature is totally indifferent to handedness, that there is a perfect symmetry between right and left handedness. One popular way of thinking of this is to try to define a right-handed object (e.g. machine screw) to an inhabitant of a distant planet or the crew of a spaceship entering the solar system. Try it! But remember that "clockwise" is a convention.

One experiment seemed to do the trick—it seemed so asymmetric that physicist Ernest Mach expressed deep shock when he thought about it. Consider a current flowing away from you. That is easy to transmit to the aliens because you can teach them how to make a battery and agree to the definition of anode and cathode (they are different elements). Now place a compass needle over the wire and it flips to the RIGHT! We can, it seems, teach the ignorant alien which is his (or her) right hand. No? No! Why not? Hint: How do you explain how to make a compass needle?

To test the symmetry, we need to discover some phenomena in which right-handedness is intrinsic. Until 1957, it was believed that this was impossible, that the world and its mirror image obeyed the same laws of physics. However, in 1957 one found a fundamental particle that had no mirror image!

Consider the muon. It has a spin = 1/2 and it is unstable. When it decays, it shoots electrons out in a variety of directions. The spin axis is a convenient reference direction —we can call it N-S or E-W or $+z$ and $-z$. We know the muon is spinning, but there is, so far, no way to distinguish one end of the spin axis from the other. We must think of this as an isolated particle. Let's do the research in outer space so that north poles, etc. do not confuse us. We have the muon (let it be a negative muon, actually a larger number of them) and a mirror and you, the observer.

Now relative to the spin axis, we can ask how the decay pattern of electrons from the demise of a large number of muons are distributed. Here is the key! If the number of electrons emitted in the decay of a number of identical, spin-aligned muons favors one end of the spin axis, say the +Z end, over the opposite end (the -Z end), then mirror symmetry is violated! This is because the lab muons, spinning say, about a vertical axis and, by our convention, counter-clockwise, favor emitting electrons upward and this defines a left-handed screw; the spin replacing the turning action of the screwdriver and the "majority of the electrons" replacing the forward motion of the screw.

In our muon example, the mirror image does not exist! Negative muons decay "left-handedly", and so, the test of mirror symmetry, or the corresponding conservation law of Parity, is the asymmetry in the distribution of electrons relative to the spin axis. In 1957, this experiment was carried out and the forward-backward ratio was 2:1. Parity would have been violated had the ratio been 1.00122 ± .00002. It turned out that the violation of mirror symmetry is a large effect but it takes place only for processes in which the weak force is involved, i.e. the decay of muons.

The question of parity (P) conservation was raised by T. D. Lee and C. N. Yang in 1956, even though this symmetry was practically bread-and-butter useful in compiling data on nuclear and atomic physics for decades. The breakthrough of Lee and Yang was the idea that symmetry could be perfectly respected in the strong and electromagnetic forces, but that the weak force, radioactivity, would ignore this symmetry. In 1957, it was discovered that weak processes are not invariant to the parity (P) operation. This was a new, revolutionary idea— that the forces of nature may have different levels of symmetry.

Charge Conjugation and Time Reversal

We have already noted that mirror symmetry, designated by the symmetry operation: reflection in a mirror, or P is not a valid symmetry when it comes to processes involving the weak force. We did discuss particle-antiparticle symmetry, which is designated by the operation C.

In 1957, this too was shown to be violated in the weak force reactions such as the decay of pions and muons. Since violations of symmetries are philosophically disturbing, they are also enticing clues to some underlying theory. After the Parity experiments showed that P is not a valid symmetry, there arose the conjecture that, perhaps, if we reflect in a mirror (P) and simultaneously change particle to antiparticle (C), we would establish the combined symmetry, PC.

Reflecting a negative muon then gives rise to a positive muon (in the mirror) and there is PC symmetry. Physicists rejoiced! We have a new and deeper symmetry which connects space and electric charge. The joy was short-lived. In 1964 in a beautiful experiment involving some more exotic particles (neutral K-mesons), it was shown that PC is also not conserved, i.e. the physics of weak forces is not invariant to the operation PC.

The origin of this breakdown of our beloved symmetry has been at the frontier of physics for the past 30 years. We still do not know how this will play out, but we have since learned that if PC were indeed a perfect symmetry, our universe would be totally different—we, our solar system, stars and galaxies, would not exist! Let this provocative conclusion close this very incomplete survey of the role of symmetry in physics.

(Go here to read in greater detail about the philosophical struggles that a deep thinker must contend with to arrive at these common sensical everyday ideas!)

Emmy Noether

Symmetry in Physics: Discrete Symmetries

Discrete Symmetries

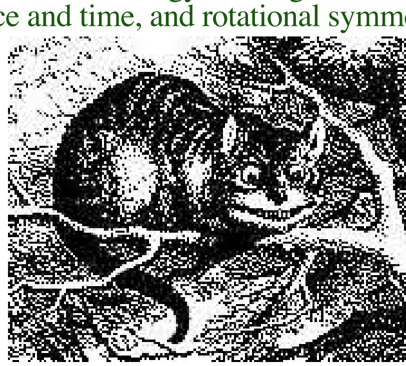
We described the way in which space and time are constructed on the basis of the continuous symmetries, i.e., the operations of translations in space, translations in time, rotations in space, and Lorentz transformations (which are essentially rotations in space-time). For each of these the symmetry operation can be an arbitrary amount, e.g., we can translate a system in space by 16 microns, or in time by 37.8 years, or rotate by 42.7^{o}, or move the system at a relative velocity of 800 km/sec. Furthermore, there is no smallest step we need take to perform a symmetry operation, i.e., we can perform an infinitesimal rotation, translation, or motion of the system. This is the meaning of the word *continuous* .

There are, however, important symmetries which are *discrete* , or non-continuous, where the symmetry operations comes in distinct steps and there is no infinitesimal symmetry operation. The operation either happens in a distinct step, or it doesn't happen at all. Three of these are of special interest, particularly when they are applied at the molecular, atomic or sub-atomic level.

Reflections in Space (Parity Symmetry)

 Essentially, this means reflecting the system in a mirror. If we look through a mirror we see another world. This we will call "Alice's world". We see physical objects in that world that move around, collide and interact, and obey a system of rules that is very similar to the system of rules that work on our side of the mirror.

For example, my cat Tum (on my side of the world) runs onto a slippery surface, a freshly waxed table top, and slides into the vase of flowers, which smash onto the floor. Momentum is conserved in this collision; energy and angular momentum are all conserved (if I include the dissipation energy in sound and heat when the flowers finally hit the floor,  the total energy is indeed conserved), etc. These are some of my laws of physics, all dicatted by symmetry principles on my side of the mirror.

In Alice's world there is also a cat, which looks very much like Tum. I'll call him Mut (he is also a "him"). He too slides on a slippery table surface and collides with a vase of flowers and knocks them onto the floor. I can actually make detailed measurements of this collision to check that in Alice's world there is exact momentum, energy and angular momentum conservation. As far as I can see, translational symmetry in space and time, and rotational symmetry, and many many other symmetries all are valid in the mirror world.  And so, I begin to believe that Alice's world, the world in the mirror is subject to the exact same laws of physics that my world is subject to.

The operation of reflecting a system in a mirror and the notion that Alice's world is governed by the same laws of physics as my world is a hypothetical symmetry. It is a discrete symmetry because we either reflect or we don't; there is no 0.126 units of reflection, it's all or nothing. This symmetry is called: Parity.

This raises an interesting and more precise question. Is parity required to be a symmetry of the laws of physics? And, indeed, is parity a true symmetry of the laws of physics? How might we find out?

Suppose that you are given a movie of a physical process. For example, it may the collision of billiard balls (or Tum colliding with the flower vase). The movie film may be flimed with the camera as shown in Fig.(1). Alternatively, it may be filmed with the camera viewing the reflection of the system in a mirror. Let's assume that this is a really good camera an a really good mirror (no knicks or smudges). Is there any way you can tell that the physical process you are viewing is filmed through the mirror, or taken directly?

Now, this is a deep question, but we need to reduce to simple systems to see that. For example, I forgot to tell you that Tum (a complex system) has a white spot on the right side of his face. Thus, Tum is "tagged" with a distinguishing feature of "right-handedness." Therefore when you view the image of the cat-vase collision you can look to see if the white spot is on the right or left side of the cat's face. If it's on the left-side, then you know you are seeing Mut, the reflection of Tum, and you can tell that the image is viewed through the mirror.

So, suppose we go down to the level of molecules. There are certain complex molecules, sugars and proteins, that come in two varieties. These are refered to as dextro-xyz (right-handed) and levo-xyz (left-handed). It is possible to tell them apart since right-handed sugars rotate the plane of incident polarized light clockwise, while left-handed sugars rotate counter-clockwise. Biological systems on earth are known to digest only right-handed proteins, while the ingestion of left-handed proteins is either useless, or fatal! Thus, the biological properties of these molecules depend not only upon the chemical (atomic) content, but also on the structure in a reflection dependent way.

If Tum had no white spot and is a completely black cat with no obvious distinguishing features, then we would have a hard time telling Mut from Tum, and we couldn't tell if the film of the cat-vase collision was shot through the mirror or not. However, we might examine the food consumed by the cat, or his DNA, or a sample of tissue, and discover that our cat had left-handed proteins of a variety that only real cats have right-handed. we could then conclude that it is Mut, and not Tum, and the film is again shot in the mirror.

By the way, the fact that all living organisms on Earth use the particular dextro-xyz form of a given sugar molecule is not surprising, and is simply a direct consequence of natural selection and evolution. In the primordial process of evolution, some 2 to 3 billion years ago, nature made a random choice to use a particular handedness molecule in a particular proto-organism. It could have easily been the other way ... it's like the kick-off of the super-Bowl, all determined by the "flip of a coin." This little proto-organism became the ancester of every subsequent living thing, and that random event therefore propagated down the chain of life to all things alive today. Evolution is a factual set of principles, and one cannot understand much of the biological world without understanding evolution.

It's not hard to make something with handedness. A box of screws in a hardware store would usually be "right-handed," i.e. a clockwise rotation moves the screw forward. Seen in the mirror, the clockwise rotation becomes counter-clockwise, but the mirror image of the screw also moves forward, i.e., the mirror screw is "left-handed." So? The point, however, is that a left-handed screw can exist in the lab, i.e., it is completely compatible with the laws of physics; nothing violates the laws of physics to make a left-handed screw, it just takes a special order from a manufacturer (e.g. "please make us 10 dozen 8-32 left-handed screws"), and so the mirror lab just has different conventions for defining left and right.

Now we go to a greater level of simplicity and we look at individual atoms in collision. Can we now tell whether the film is taken through the mirror or not? Even atomic and nuclear collisions fail to reveal any difference between a given system and its mirror image. Physicists thought that once we got down to simple systems, systems that are not constructed from a complicated set of rules such as a cat (involving natural selection, and many stages of evolution in which handedness does become imprinted), that we would see pure left-right symmetric laws of nature. At this level we shouldn't be able to tell if the film is taken through the mirror of otherwise.

Nonetheless, we go still deeper and test this idea at the level of elementary particles. Are there any properties of elementary particles that are different in Alice's world from our world?

We'll describe an experiment done by one of us (Leon Lederman) to test the idea of parity as a symmetry in the interactions of elementary particles. It provides a fairly simple way of seeing what's going on, but you will have to think it through carefully. Here goes:

There is an elementary particle called the pi-minus meson, denoted π^- (which, we now know today is made of a down-quark and an anti-up quark, so it really isn't an "elementary particle" at all, but for our purposes we can consider it to be such) The π^- decays into a muon μ and a neutrino ν^0, or:

$\pi^- \rightarrow \mu^- + \nu^0$,

The π^- meson is spin-0, meaning it has no angular momentum; it is a sort of tiny blob which does not rotate. The muon and the neutrino on the other hand each carry spin; they are little pin-points with angular momentum. We say they have spin-1/2 because the quantity of angular momentum which they carry is precisely 1/2 times Planck's constant.

Now, we know, according to Noether's theorem and rotational symmetry, that the conservation law of angular momentum must hold, even for tiny elementary particles. Thus, when a pi meson decays, the initial angular momentum is zero, and therefore the sum of the angular momenta of the muon and the neutrino must be zero. An extremely important experimental point is that you can slow down and stop a speeding muon, and you can measure its spin. The slowing down and stopping doesn't change the spin direction, so it is actually possible to measure the exact direction of the muon angular momentum!

So, we can look for events where the muon comes out with its spin aligned along the direction of motion. And we can look for events in which the spin is counteraligned to the direction of motion. When the spin is aligned in the direction of motion we say the helicity of the particle is positive (+); when the spin is counteraligned to the direction of motion we say the helicity is (-).

Now, you must think the following amazing fact through carefully: Helicities are always reversed when viewed in a mirror.

Recall how we define the angular momentum vector using the righthand rule. For a top, we curl our right-hand fingers in the direction that the material is spinning, and our thumb defines the angular momentum vector direction. (See the figure) This is a convention that we use, and it must be used consistently for everything, e.g., we use the right-hand rule for muons and for neutrinos; we don't switch the left-hand rule when we switch between muons and neutrnos. Also, since we don't know a priori if we are viewing a movie filmed through a mirror or not, we will always use the right-hand rule, even for systems we see in Alice's world, i.e., we don't switch to a left-hand rule in Alice's world, because there is no way of knowing if we are viewing Alice's world or our world. Now consider a tennis ball that is spinning and moving in any direction, with spin aligned in the direction of motion (recall, we use the right-hand rule to construct spin). It's mirror image may have the direction of motion reversed, but then the spin will not reverse (using right-hand rule for the mirror image!); or the direction of motion may be the same, but then the spin-direction reverses. So, the helicity is always reversed in the mirror. This is also true of screws, if you think in terms of the helicity of a screw, Also, think of the mirror image of a winding staircase and you will realise that helicity is reversed there as well.

Leon measured the helicity of the outgoing muon in pi decay. If parity is a good symmetry of the laws of physics, then both helicity + and helicity - should occur with equal probability (remember, quantum theory produces things probabilistically). In fact, the result of the experiment was amazing: The helicity of the muon proucude in π^- decay is always (-).

Why is this so amazing? Because, if you were to "see" a film of a pion decay into a helicity (+) muon, then you would be able to announce instantly that this process is occuring in Alice's world; that is, you are seeing an image through a mirror!!! The laws of physics contain interactions which are not symmetric under parity, and these intereactions occur in pi meson decay. This is called the left-handed neutrino (or right-handed anti-neutrino) and it is produced in many processes called weak interactions. The spin of the neutrino is always counteraligned to the direction of motion. Hence the term "left-handed." Neutrinos are so weakly interacting that they can travel unimpeded through vast distances of solid matter. In fact, there are so many neutrinos being produced in the sun that billions of them are passing through you every second.

Until 1957, it was believed that parity was an exact symmetry of physics. The question of parity (P) conservation was first raised by T. D. Lee and C. N. Yang in 1956, even though this symmetry was practically bread-and-butter useful in compiling data on nuclear and atomic physics for decades. The breakthrough of Lee and Yang was the idea that symmetry could be perfectly respected in the strong and electromagnetic forces, but that the weak force, radioactivity, would ignore this symmetry. In 1957, it was discovered experimentally that weak processes are not invariant to the parity (P) operation. This was a new, revolutionary idea— that the forces of nature may have different levels of symmetry.

Reflections in Time (Time Reversal Symmetry)

This symmetry examines the effect on the laws of physics of reversing the direction of the flow of time. Newton's laws of motion make no distinction between past and future, and time can apparently flow in any direction.

Again, we can view the laws of physics in action in a movie film. We then run the film backwards through the projector. When applied to simple systems, billiard balls colliding on the table, atomic collisions, etc., it would not be possible to tell in which direction the film was progressing. The motion we see satisfies laws of motion that are the same, whether run forward or backward.

However, when we apply this to more complex objects, like physicists, it is easy to tell which way the film is running -- their hair always turns grey, the total useful reserve of energy alwasy decreases as we go into the future, and Humpty-Dumpty never reassembles himself and jumps back up on the wall. The motion of large complex (or statistical) systems including people, is not time reversal symmetric, even though the simple components obey time reversal symmetric laws of motion.

Notice that in physics we always pose and solve if-then problems. For example, consider the following question (Q1): If a particle at time t_1 is located at x_1 traveling at a velocity, v, then where will it be at time t_2? Of course, the answer is trivial:

$$x_2 = x_1 + v(t_2 - t_1)$$

but even this trivial result illustrates some deep philosophical issues as to how we describe nature. First, we see indeed that the answer is time reversal invariant. A time reversed question is (Q2): "If at time t_1 the particle is located at x_2 and traveling with velocity -v (velocities change sign when we reverse the arrow of time), then where will it be at time t_2?" Now the answer must be x_1. And indeed, we see that our formula gives:

$$x_1 = x_2 - v(t_2 - t_1)$$

In this way of time-reversing the question, we set up initial conditions that were the opposite to those in the first question Q1, i.e., we put the particle at the location where it ended up in Q1, and we reversed the direction of motion. We thus find that after an equivalent time interval, the particle gets back to x_1. So, we can do time reversed physics without actually reversing the flow of time.

Notice another peculiar aspect of physics. No where in any formulation does the issue of a special point in time called "now" ever occur. Yet, we humans sense something we call "now." Is it an illusion? We call this the "Now" question.

Relativity tells us that there is no absolute "now" in the Universe, because different observers in different inertial reference frames will disagree on which events at different places in space are simultaneous. Hence, even within our brains, there can be no perfect synchronization on short time scale of order (size of brain)/(speed of light). However, the brain is fairly slow, so perhaps there is some averaging going on that yields the experience of "now." Is "now" therefore real and part of the laws of physics?

The fact that this question is so murky tells us the answer: There is no priveleged role of "now" in the laws of physics. The perception of the sensation of "now" has to do with the murky business of "consciousness." Since there is no real theory, or even model, of consciousness, we cannot address this further, except to say that consciousness is a very complex phenomenon.

So, why do complex systems seem to prefer an arrow of time, while their elementary constituents do not? This also has to do with the if-then nature of physics. Anything we observe involves laws of motion, but also special initial conditions, called "boundary conditions" in the theory of differential equations (the laws of physics are, for the most part, differential equations). If I have initially a container full of a gas, and I open the valve on the container, the gas will escape and fill a room. The laws of motion are perfectly time reversal invariant, bt I never see a room full of gas collect itself spontaneously into a container. It is simply very unlikely to have an initial condition of a gazillion gas molecules with their velocities and positions at the initial t_1 such that they will collect into the bottle. We can introduce a statistical concept, a measure of randomness, called entropy (we won't bore you with the precise mathematical definition of entropy). In perfect equilibrium, like hot soup sitting in a thermos bottle with no escaping heat, the entropy remains constant in time; in non-equilibrium processes, like shattering glass, or rotting flesh, the entropy always increases.

Now, this does not mean that complex ordered systems cannot evolve. Indeed, they certainly can and do evolve. As a system, such as a gas of water vapor cools, there will form droplets of liquid, which have more order than the gas; still cooler and the droplest form crystals of ice with still more definite order. This process of cooling is not an equilibrium situation; energy has been allowed to flow out of the water vapor (perhaps as radiation, i.e., photons). As the radiation scatters out into space, occupying a more chaotic distribution (more entropy), a small subsystem of cooled droplets is left behind (less entropy). The overall entropy has increased, while a subsystem seems to have formed with less entropy. If that subsystem contains a certain configuration of molecules, such as a nucleic acid, then it may be able to make copies of itself by expending more energy out into space; again overall enetropy increases, but we get a more and more complex subsystem left behind. And eventually, we can form a human sitting there wondering why time seems to flow in a particular direction.

The complex subsystem, (if) having been formed, (then) can evolve in a way by which it increases its own entropy: It can fall apart, rot, dissolve, or fade away.

Particle-Antiparticle Symmetry (Charge Conjugation)

A discrete symmetry of replacing all particles by anti-particles in any given reaction is called C. The symmetry would imply that the laws of physics are exactly the same in the anti-particle world as they are in the world. For example, anti-Hydrogen consisting of an anti-proton and an anti-electron (positron) would have the same properties, e.g., energy levels, sizes of the electron orbitals, etc., as does the ordinary Hydrogen atom (by the way, anti-Hydrogen has actually been made in the laboratory at CERN and Fermilab in the past few years).

We have already noted that mirror symmetry, designated by the symmetry operation: reflection in a mirror, or P is not a valid symmetry when it comes to processes involving the weak force. The discrete symmetry of replacing all particles by anti-particles is called C. And, as we have seen, there exists yet another discrete symmetry operation, called T, which reverses the flow of time, i.e., set t -> -t in all physics equations.

In 1957, C was shown to be violated in the weak force reactions such as the decay of pions and muons.

Since violations of symmetries are philosophically disturbing, they are also enticing clues to some underlying theory. After the Parity experiments showed that P is not a valid symmetry, there arose the conjecture that, perhaps, if we reflect in a mirror (P) and simultaneously change particle to antiparticle (C), we would have the combined symmetry operation, CP, and perhaps this symmetry is exact in nature.

Doing CP to a negatively charged muon then gives rise to a positively charged muon (antiparticle, in the mirror) and this turns out to be a symmetry of the muon decay! Physicists rejoiced! We seemed to have a new and deeper symmetry which connects space and electric charge.

The joy was short-lived, however. In 1964, in a beautiful and extremely well executed experiment, involving some more interesting mesons (called neutral K-mesons, each containing a strange and anti-down, or a down and anti-strange quark), it was shown that CP is also not conserved, i.e. the physics of weak forces is not invariant to the operation CP.

The details of the origin of this breakdown of the symmetry CP has been at the frontier of physics for the past 30 years. We still do not know how this will play out, but we have since learned that if CP were indeed a perfect symmetry, our universe would be totally different and we, our solar system, stars and galaxies, would not exist!

Now, it is a necessary condition in quantum mechanics that the combined operations of CPT must be an exact symmetry. It turns out that if CPT failed as a symmetry, then over time probablity would not be conserved; i.e., the probability for anything to happen under any circumstances would either exceed or be less than one! Nevertheless, we ask, if the violation of CPT were very, very tiny, would we have noticed? It is, afterall, an experimental question. And it is a deep theoretical question when one considers mini-black-holes forming and disappering at very small distances in the vacuum. Do black-holes eat probability?

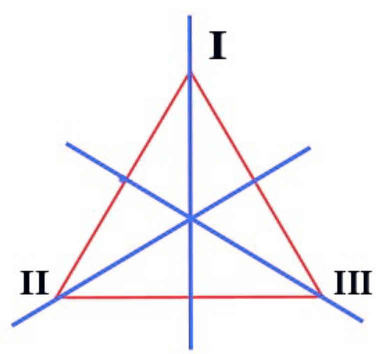
Recently, in refined experiments with neutral K-mesons, direct confirmation of the violation of T-symmetry has been confirmed. T must be violated when CP is violated in such a way as to make CPT conserved. Hence, to date, we expect CPT is an exact symmetry. However, T-violation means that deep down in the laws of physics there is fundamental information that defines a preferred direction, or arrow, of time. Time reversal invariance is not a symmetry of the laws of physics.

Emmy Noether

Symmetry in Physics: Mathematics of Symmetry

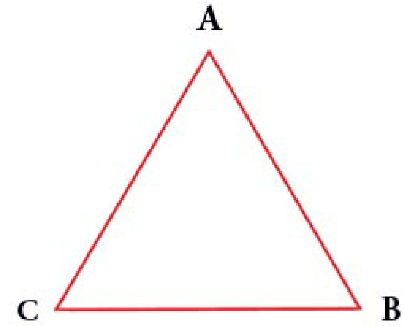
Mathematicians solve many problems in geometry and topology by turning them into *equivalent algebraic problems*. This approach to understanding symmetry begins with the 19th century French mathematician, Galois, who in his short, tragic life laid the foundation of what we call "group theory".

Let us think concretely about the symmetries of a very simple geometric object... the equilateral triangle. This is the simplest nontrivial example of a system possessing symmetry and the first time through you may find the mathematics hidden in this simple object to be quite surprising!



It is useful to approach this experimentally (if you can visualize the following manipulations, that's fine; but you are strongly encouraged to perform the experiments yourself). Prepare two transparencies or cut-outs as in Fig.(1) and Fig.(2) each featuring an equilateral triangle, both of the same size. The triangle of Fig.(1) has the three axes of symmetry labeled as I, II and III. The triangle of Fig.(2) has the vertices labeled as A, B, and C. (You can just print these figures onto transparencies, or onto paper and then copy them onto transparencies, or whatever high-tech solution you prefer; we prefer to draw them onto blank transparencies with colored pens).

The triangle (1) is laid down on a table (or in classroom, onto a transparancy projector platten) and should be considered to be "glued" or "taped" in place. It is our *reference triangle* and it serves as a reference grid, or a kind of "coordinate system". Once laid in place we will not move it again. Triangle (2), on the other hand, is an *experimental triangle*. We will overlay the reference triangle with the experimental triangle. Our problem is to *find all possible distinguishable ways in which the experimental triangle can be lifted up and brought down on top of the reference triangle* . The vertices of the experimental triangle are labeled so as to allow us to identify the distinguishable ways in which this can be done.



We shall begin by overlaying the experimental triangle on the reference triangle with the vertices reading ABC clockwise around the experimental triangle. This will be called the *initial position*. We wish to find a way in which we can pick up the experimental triangle and bring it back down on top of the reference triangle so that the vertices read something other than ABC clockwise. Each such operation is called a *symmetry operation* . Our problem is to dicover all possible distinguishable symmetry operations of the equilateral triangle. How do we proceed?

Perhaps you've considered rotating the experimental triangle until the vertices now read CAB clockwise from the top. This certainly corresponds to a symmetry operation, which is a rotation through 120^o. We shall designate this first discovery as R_{120}.

Now return to the initial position. What else might we do? It is obvious that a rotation through 240^o is another symmetry operation which yields the result BCA. However, it is important to emphasize that we should always return to the initial position before performing the next operation (this is a bit like pressing the *CLEAR* button on a pocket calculator before doing the next calculation). Thus we discover a second distinguishable symmetry operation which we designate R_{240}.

Are there other symmetry operations? Perhaps you've considered a rotation by -120^o. However, we now see that this operation takes the triangle to BCA (from the initial position, of course) and therefore this is not a new operation, i.e. it is not *distinguishable* . (We may thus write the equation: $R_{-120} = R_{240}$)

What about a rotation through 360^o? First, we see that it maps the triangle from the initial position ABC back to the initial position ABC. Consequently, it *is* a symmetry operation, but a very special one. For one, it is equivalent to doing nothing at all; as such we shall refer to it as the "do nothing operation", or *the identity operation*. We shall denote it by the boldface $\mathbf{1}$. Second, note that the identity element is a symmetry operation of any object; even a lowly Amoeba has the identity symmetry. Thirdly, we note that a rotation through 360^o is equivalent to a rotation through any *integer multiple* of 360^o, e.g., 720^o, 1080^o -360^o, etc. All are equivalent to the "do nothing" operation, $\mathbf{1}$.

We now have three symmetry operations; are there more? We may consider a *reflection* about one of the three axes of the reference triangle. We begin with the initial position and consider "skewering" the experimental triangle (as if we had a barbecue skewer) along one of the axes of symmetry. For example, skewering along axis one, we then pick up the triangle and flip it and we arrive at the new position, ACB. We denote this symmetry operation as a *reflection about axis I* and, we'll give it a symbolic name as well: R_I.

Similarly, we return to the initial position and consider the other two reflections (a) about axis II, or R_{II} which yields the position BAC and (b) about axis III, or R_{III} which yields the position CBA.

Thus at this point we have a list of six symmetry operations:

1 *identity* "do nothing" ... (ABC)

R_{120} *rotate by* 120^o ... (CAB)

R_{240} *rotate by* 240^o ... (BCA)

R_I *reflect about axis I* ... (ACB)

R_{II} *reflect about axis II* ... (BAC)

R_{III} *reflect about axis III* ... (CBA)

Are there any other symmetry operations? At this point perhaps you've recognized that we have discovered essentially the six permutations of three objects, 3!=6, i.e. the six permutations of the three vertices of the triangle.

That raises an interesting question:

Q: Are the symmetries of all such objects, such as squares, pentagons hexagons, cubes, etc. given by the permutations of their vertices?

A: In fact the answer is *no* . It doesn't work that way for the square as we can easily see. Suppose we have a square with vertices labeled ABCD. A typical valid symmetry operation of the square is a rotation through 90^o and gives DABC, which is, indeed, also a permutation of the vertices. However, is there a symmetry operation which can give the vertex ordering BACD? (Think in terms of an experimental square on a cut-out; what would we have to do to the cut--out to get BACD starting from ABCD?) Clearly, this is not a symmetry operation of the entire square because we would have to twist the experimental square to get the vertices into this position, but then the sides would not overlay properly! Thus, while all symmetry operations are indeed permutations, not all permutations are symmetry operations. The equilateral triangle is simpler and it does have only six symmetry operations, the ones we've listed above, which are equivalent (isomorphic) to the permutations of three objects.

Another question may be bothering you:

Q: Why don't we distinguish between symmetry operations like R_{120} and R_{480} ? What happens if we try to distinguish between these?

A: Good question. In fact, if you try to distinguish between operations like the identity and R_{360}, or between R_{120} and R_{480}, then you are really not focused on the intrinsic symmetry of the triangle. Instead, you would be focused more on the *path that we take* when we perform the symmetry operation. For example, I could perform an operation like R_{120} by picking up the experimental triangle and placing it back down rotated by 120^o in the usual way. Or I could do it by picking up the experimental triangle, and then go outside and run around a tree in my back yard 10 times, and then come back in the house, eat a doughnut, and then place it back down rotated by 120^o. There is no added content to the analysis of symmetry in adding all of this running around to the operation. Indeed, I won't even know if I am doing R_{120} or R_{480} or $R_{120\,+\,360N}$ where N is a large integer (possibly negative) if I do it the second way. So, we find that the essence of maximal distinguishability is captured in the six symmetry operations on our list. Incidently, there are other branches of mathematics, such as homotopy, that are interested in the *paths we can take* on different surfaces, or in different spaces.

Thus far our exercise has been almost trivial, but now we make the great observation of Galois and his colleagues. We now ask, can we obtain additional symmetry operations by combining together two of the operations previously obtained? That is, let us select any two of our six operations, say R_{120} and R_{II}. Let us first perform one of the operations on the experimental triangle (try R_{120}) and *without returning to the initial position perform the other operation* R_{II}. We see that if we begin in the initial position that R_{120} leads to CAB and then following with R_{II} we obtain the position ACB. But ACB is not a new position of the triangle, and we see from the above list that it corresponds to R_I. We have therefore discovered an interesting result: first performing R_{120} and following it by R_{II} yields the result R_I .

Let us write an equation for this result:

$$R_{120} \text{ x } R_{II} = R_I \,.$$

Here we have introduced a symbol, x, which represents the action of combining the symmetry operations in the order indicated. It is easily seen that the x combination of any pair of our symmetry operations (which we also refer to as "elements") produces another of the elements. We say that our set of elements is *closed* under the operation x. Thus, in a sense the combining of two symmetry operations is something like *multiplication of numbers* . In this sense the "do nothing operation" is the true identity since:

$$1 \text{ x } R = R \text{ x } 1 = R \text{, for any operation } R \,.$$

	1	R_{120}	R_{240}	R_I	R_{II}	R_{III}
1	1	R_{120}	R_{240}	R_I	R_{II}	R_{III}
R_{120}	R_{120}	R_{240}	1	R_{III}	R_I	R_{II}
R_{240}	R_{240}	1	R_{120}	R_{II}	R_{III}	R_I
R_I	R_I	R_{II}	R_{III}	1	R_{120}	R_{240}
R_{II}	R_{II}	R_{III}	R_I	R_{240}	1	R_{120}
R_{III}	R_{III}	R_I	R_{II}	R_{120}	R_{240}	1

Thus we have made a very important observation: *the symmetry operations form an algebraic system with an operation consisting of performing successive operations* . This algebraic system is called a Group , or a Symmetry Group . The symmetry operations are the analogues of the rational numbers under the group multiplication. We refer to the symmetry operations as "group elements" or simply as "elements" .

We present the complete multiplication table of the symmetry group of the equilateral triangle in the table to the right. You should verify this for yourself by performing several of the cases with the experimental triangle and reference triangle transparencies. The table is to be read like a highway mileage map; if we choose to perform the product R X R' we first find the row labeled by R , then the column labeled by R' and we look up the corresponding entry in that row and column. Take a moment to study the group multiplication table. Notice, for example, that every element of the group occurs once and only once in every row, and in every column of the table! The multiplication table of symmetry operations is a magic square!!!

There are several important properties of all symmetry groups. In fact, this is the precise mathematical definition of a group:

A group is a set of elements and a composition law, x, such that the product of any two elements yields another element in the set, (closure);

Every group has a unique identity element satisfying $1 \times R = R \times 1 = R$ for any element R of the group.

Each element of the group has a unique inverse element. That is, given an element R there exists one and only one element R^{-1} (which may even be R itself), such that $R \times R^{-1} = R^{-1} \times R = 1$.

Group multiplication is associative.

Associativity is a bit tricky. It means, given any three group elements R_1, R_2 and R_3, then:

$R_1 \times (R_2 \times R_3) = (R_1 \times R_2) \times R_3$

In words, start with the triangle in the initial position. First perform operation R_2 and follow it by R_3 and remember the result (call this result R_4). Now return to the initial position of the triangle and first do R_1 and follow by R_4. The result of this sequence of operations will always be the same as having first done R_1 followed by R_2 then followed by R_3. This seems complcated, but is the true operational meaning of *associativity* and you should carefully think it through to make sure you understand it. Indeed, we take this for granted because the ordinary operations of arithmetic are associative, i.e. 3x(4x5) = (3x4)x5. However, there do exist, in pure mathematics, nonassociative systems in which Ax(BxC) does not equal (AxB)xC. These include "normed division algebras based upon octonions." People have attempted to relate this kind of mathematics to physics (octonions were thought to be possibly associated with the physics of quarks in the mid 1970's), but it seems not to be relevant. Thus, as far as we can tell, nature and its associated mathematical description is always associative.

It is remarkable that these definitions (or axioms) that define groups in a logical or algebraic sense capture the essence of symmetry. All symmetries are groups, and groups always have a geometric interpretation as a symmetry (though it can get to be a very complicated geometrical interpretation; for example, the "monster" group corresponds to the symmetry of the closest packing of ball bearings in a crate that lives in 26 dimensions!)

From the "axioms" on our list above one can prove some very important theorems:

Each element of the group occurs once and only once in each row and each column of the multiplication table. This can be proved as a theorem from the preceding statements.

This is a powerful constraint on the mathematical structure of the group; essentially the group multiplication table forms a kind of "magic square".

Here is a real thriller: Group multiplication is not necessarily commutative! That is, (R_1 x R_2) need not equal (R_2 x R_1)!!!

This last result, namely that group multiplication is not commutative, is really quite remarkable. We can see it by doing an example. Start in the initial position of the triangle and first perform R_{240} and then follow by R_{II}, i.e. calculate R_{240} x R_{II} You should obtain the result R_{III}. On the other hand, return to the inial position and now first performing R_{II}, followed by R_{240}. The result now is R_I. Summarizing:

R_{240} x R_{II} = R_{III}

R_{II} x R_{240} = R_I

This is truly remarkable! Here we have discovered a simple system of six elements with a multiplication law and the system is not even commutative. Thus, although the ordinary multiplication we learned in grade school is commutative, e.g. 3x4 = 4x3, group multiplication need not be (Do you remember being bored silly by the notion of commutativity in grade school? It seemed like an empty and trivial statement at the time; of course multiplication is commutative. Yet, we see now that with the marvelous algebra of groups this ceases to be the case. Commutative algebras are a very special case indeed). Now, when a group has completely commutative multiplication we give it a special name: it is said to be an *abelian group* , after the mathematician Abel. The general group, such as the equilateral triangle group, is noncommutative, or *nonabelian* .

All possible continuous groups, i.e. groups with an infinite number of operations that vary continuously with "angle" parameters, like the rotations of a sphere about a given axis through any angle, were completely classified early in the 20th century by Cartan. Remarkably, only very recently have all possible discrete symmetry groups been classified. This job was made difficult by the existence of certain "sporadic" groups, such as the "monster group" with about 8x 10^{53} elements. The classification of the discrete groups constitutes one of the longest and least comprehensible theorems in mathematics (see "The Enormous Theorem," by D. Gorenstein, *Scientific American* (Dec. 1985), pg. 104).

Group mathematics seems to underlie the structure of our physical world. One may wonder how a noncommutative mathematics can have anything at all to do with nature, or physics? Yet it is easily demonstrated in the classroom. For example, perform two successive rotations by 90^o on a textbook, first about an imaginary x-axis followed by one about an imaginary y-axis, and note the book's position (remember also your choice of x-axis and y-axis, and don't change these!). Now return the book to the initial position and perform first the rotation about the y-axis followed by a rotation about the x-axis. You will find that the book ends up in a different position the second time than you obtained the first time. In fact, the rotations can be any angles you choose, and the same noncommutativity will occur. Thus, our real world in 3-dimensions is very very noncommutative!

The symmetry group consisting of rotations of objects through any angle about three chosen orthogonal axes is space is a noncommutative group. Indeed, this group has a name: The continuous group consisting of all rotations of objects in three dimensions (the full symmetry group of a sphere) is known as SO(3). It governs the physics of angular momentum and spin.

We can generalize this group of rotations. The symmetry of a sphere in N-dimesnions is called SO(N). Mathematically, we consider an (unit length) N-vector which lives in an N-dimensional space. We can describe this N-vector by giving its N components $(x_1, x_2, \ldots , x_N)$, i.e., these components are just the projections of the vector onto some orthogonal N-dimesnional coordinate system. Now the action of the symmetry group SO(N) is to replace these N-components by new values, $(x'_1, x'_2, \ldots , x'_N)$. The condition is that both the x_i and the x'_i must be unti vectors, i.e., they describe points on the surface of a unit-sphere (a sphere of radius = 1) in N-dimensions, i.e.,

$$x_1^2 + x_2^2 + \ldots + x_N^2 = 1$$

$$x'_1{}^2 + x'_2{}^2 + \ldots + x'_N{}^2 = 1$$

Yet another generalization produces very important symmetry groups in quantum theory. Now we view our N-vector as a set of N complex numbers: $(z_1, z_2, \ldots , z_N)$. The group SU(N) replaces these by new numbers $(z'_1, z'_2, \ldots , z'_N)$ such that:

$$|z_1|^2 + |z_2|^2 + \ldots + |z_N|^2 = 1$$

$$|z'_1|^2 + |z'_2|^2 + \ldots + |z'_N|^2 = 1$$

Hence, SU(N) is the symmetry of the complex unit sphere. Complex numbers play a fundamental role in quantum mechanics, and thus the usual symmetries occuring in nature are SU(N) symmetries.

In the mid 1960's it was recognized that the strongly interacting particles (like protons, neutrons, pi-mesons, and particles called "strange" particles etc.) could be placed into multiplets of a continuous symmetry group, SU(3). (see "The Eight--Fold Way", by M. Gell-Mann and Y. Neeman, (1964).) One of the representations of SU(3) has eight components, and is known as an octet (these SU(3) symmetry elements can be represented as 8x8 matrices in an irreducible way; the eight members of the octet mix amongst themselves under an SU(3) transformation). Eight known spin-0 mesons fit into one multiplet, the eight spin-1/2 baryons into another, and so on. There is also a 10 component representation into which the spin-3/2 baryonic resonances fit (in fact, one, the Omega-minus, was missing at the time SU(3) symmetry was discovered and it was correctly and dramatically predicted by Murray Gell-Mann in 1963). The particles in the multiplets were not degenerate (having equal masses) indicating that the SU(3) symmetry was not exact, but the pattern was clearly established.

The main puzzle was then that the smallest representation of SU(3), namely a triplet consisting of three spin-1/2 particles with predicted electric charges of +2/3 (up), -1/3 (down) and -1/3 (strange) were not detected in experiments (in these units the electron charge is -1). These particles are known respectively as the "up quark", the "down quark," and the "strange quark". Today, however, with more powerful microscopes (i.e., particle accelerators and particle detectors) we have seen the quarks living deep inside of the all the strongly interacting particles. Indeed, the list of quarks now contains three additional ones, "charm," "bottom" (aka "beauty"), and "top" (aka "truth"). All of the mesons produced in high energy collisions are composed of quark and anti-quark combinations, while each baryon contains three quarks. We now know that the forces between quarks permanently lock them inside of these specific combinations. So, while the quarks are like the basic "atoms" in nature, we only find them inhabiting combinations or "molecules;" we have never seen an isolated quark in any experiments performed to date.

Thus, symmetries have fancy names given by the mathematicians. The spherical symmetry is called O(3); the cylindrical symmetry is called O(2). Now, notice that any symmetry operation that we can perform upon the cylinder we can also perform upon the sphere; but the converse is not true.

The symmetry group O(2) (of the circle) fits inside of the symmetry group O(3) (of the sphere). You can think of it this way: Suppose you made a perfect sphere out of silly putty. It would possess the symmetry operations of O(3). Now flatten the sphere into a pancake. This pancake is just a slightly fat circle and it therefore has the symmetry group O(2). You can still perform the symmetry operations of rotations about the pancake axis of symmetry, but we have lost all of the other operations that were the (larger) symmetry group of the sphere. Hence we say that O(2) is a "subgroup" of O(3).

We can give a more precise mathematical definition of the continuous group symmetry operations as follows. We'll consider the group O(2). Consider the pair of real numbers (x,y). The special pairs that satisfy $x^2 + y^2 = 1$ define the unit circle. Now consider a linear mapping from any pair to a new pair (x,y) -> (x',y') such that $x'^2 + y'^2 = 1$. This mapping is a symmetry of the circle. In fact, the term "linear" is all important here because there are an infinite number of nonlinear maps of the circle into itself (some which squeeze and stretch the circle perimeter, maintaining the circle's shape). The symmetry group of O(2) in its simplest form is considered to be the linear maps (x,y) -> (x',y') such that $x'^2 + y'^2 = 1$, and therefore we can write them down:

$x' = x \cos(\theta) + y \sin(\theta)$

and

$y' = y \cos(\theta) - x \sin(\theta)$.

Similarly, we can consider the sphere in three dimensions as the set of points (x,y,z) satisfying $x^2 + y^2 + z^2 = 1$. Then the set of linear maps (x,y,z) -> (x',y',z') such that $x'^2 + y'^2 + z'^2 = 1$ defines the symmetry group O(3), as we described above.

Another way to visualize O(2) is to consider the complex plane, i.e. the ordinary plane of ordered pairs (x,y), where we represent each point by a vector z = x+iy. z is a complex number; i is the square root of -1. Given any complex number z we can obtain a new complex number z' by multiplying z by the complex phase factor $\exp(i\theta)$, that is:

$z' = \exp(i\theta)\, z$

where θ is real. Its easy to see that z' is just the vector z rotated counter-clockwise by the angle θ

Therefore, if we consider the set of z such that $|z|=1$, we have the unit circle. The transformation $z' = \exp(i\,\theta)z$ just maps the circle into itself. Hence, the symmetry group of the circle is just the set of all operations, $\exp(i\,\theta)$. We see that the set contains the identity, and that the inverse of any rotation $\exp(i\theta)$ is just $\exp(-i\theta)$. [Actually, we restrict θ to lie in the interval $2\pi > \theta >= 0$ so that there is a one-to-one-correspondence between rotations and the parameter θ].

The group of rotations on the complex numbers is called the one-dimensional unitary group and is denoted U(1). However, we see that it is completely equivalent to the group O(2). O(2) is generally considered to act on real 2-d vectors, but complex numbers can be used to represent such 2-d vectors, and hence the equivalence between the two groups.

The Homepage.

I. Introduction.

II. What is Symmetry?

III. Symmetries of Space and Time.

IV. Special Relativity

V. Noether's Theorem and Conservation Laws.

VI. Conservation Laws and Quantum Mechanics.

Emmy Noether

The Problem of Mass

More than 60 years ago Enrico Fermi scribbled down a descriptive theory of the "weak interactions." These were feeble forces seen at the time at work in nuclear processes. Fermi had to introduce a new fundamental constant into physics, which we call G_F (the F stands for Fermi). This fundamental constant contains a fundamental unit of mass, which sets the scale of the weak forces, and is about 175 GeV. (= about 175 times the proton mass; 1 GeV = 1 Giga electron volts; we use energy to describe mass because $E = mc^2$; the proton has a mass of approximately 1 GeV).

In the intervening years we have come to understand a great deal about the weak forces. In the early 1970's the greatest stride along this path occured when the "Standard Model" was theoretically and experimentally established. This is a true unified theory of weak, electromagnetic, strong and gravitational forces under one fundamental symmetry principle, called "the gauge principle." Like the discovery of DNA as the basic information carrier of all living things, the gauge principle is the basic underlying defining concept of all forces we know of in nature. Yet, despite this triumph, the origin of scale of the weak forces as embodied in Fermi's original theory, the 175 GeV, remains a subtle mystery.

In the Standard Model we view the vacuum as something like a superconductor. In a superconductor, which can be made in the laboratory (they are used in many commercial devices, such as medical magnetic imaging systems, sensitive magnetometers, etc.) a quantum effect at very low temperatures causes the photon to become a massive particle in a material such as Lead or Niobium. In free space the photon is perfectly massless (hence, it always travels at the speed of light). Yet, in a superconductor a photon becomes heavy, with a mass of about 1 electron volt, and, in principle, it can be brought to rest. This mass generation for the photon gives rise to the peculiar features of superconductors, e.g., they have absolutely no electrical resistance to current flow.

In the Standard Model, however, it is *not* the photon which becomes massive (the photon remains massless in the Standard Model!). Rather, 3 other particles, closely related to the photon, the W^+, W^- and Z^0, become heavy. *In fact, the forces that are mediated by the quantum exchange of W's between other particles are exactly the weak forces that Fermi was trying to understand 60 years ago* . Indeed, the weak forces are weak because the W and Z are heavy. We conceive of the same kind of quantum effect as in a superconductor acting everywhere throughout the Universe to give the W's and Z their masses, in analogy to the way that the superconductor gives the photon a mass. The Fermi scale is the direct quantitative measure of this phenomenon. What then is this new quantum effect that gives rise to mass in the Standard Model?

Taking our cue from the superconductor, we find that "something" must undergo "condensation" in the vacuum itself (this is analogous to Bose-Einstein condensation, and it also occurs in "superfluids"). Condensation is only vaguely something like the formation of dew on a lawn in the cold early fall. In quantum theory a condensate is a very large value of a quantum field filling all (or much) of space. This "something" can be modeled by a spin-zero field that fills all of space and is called the "Higgs field," after Peter Higgs of the University of Edinburgh. The strength of the Higgs field in the vacuum is measured as an energy and *it is just the Fermi scale,* 175 GeV!

The condensate is felt by the various particles through their "coupling strengths." For example, the electron has a coupling strength g_e. Then, the electron mass is determined to be: $m_e = (g_e) \times (175\ \text{GeV})$. Since we know $m_e = 0.0005$ GeV, we see that $g_e = 0.0005/175 = 0.0000029$. This is a very feeble coupling strength, so the electron is very low mass particle. Other particles, like the top quark which has a mass $m_{top} = 175$ GeV has a coupling strength almost identically equal to one. Still other particles, like neutrinos, have nearly zero masses, and therefore nearly zero coupling strengths (the topic of neutrino masses has recently become very interesting, but we won't delve into it here).

The Standard Model only predicts the coupling strength of the W^{+-} and Z^0 particles, so their masses, M_W and M_Z (note that W^+ and W^- are just particle and antiparticle, and must have the same identical masses; the Z^0 is it's own antiparticle), are predicted (correctly) by the theory. The W has a mass of about 80 GeV and the Z a mass of about 90 GeV and these are presently being measured to very high precision. These particles, W and Z, called "gauge particles," have coupling strengths that are related to known quantities, like the electric charge, e, and something we measure in the lab called the weak mixing angle, θ_W. So, knowing e, θ_W and G_F, we can predict M_W and M_Z.

So what is the Higgs field? We don't know. It is put into the Standard Model "by hand" to explain mass. We know there must be something there that either really is a Higgs field, or that imitates one in a very faithful way. We do know that we can, if something like a Higgs really exists, with sufficient energy and luminosity, give the vacuum a good kick and produce a Higgs-like particle in the laboratory. This particle is believed to have a mass less than 200 GeV (if it really is a pure Higgs), and may even be lighter than that. As you may know, the LEP machine at CERN, which has reached the end of it's lifespan and is now being decommissioned, was producing some faint hints of a possible Higgs boson signal at 115 GeV (many people wanted to continue to run to find out if these hints were real, but CERN must now begin construction of the very high energy LHC).

At Fermilab, over the next 5 years, we have a very good chance of producing and observing the Higgs. To understand this, first you must understand that the Tevatron is essentially a quark collider, in fact most of the interesting physics is produced at the Tevatron when an up or down quark inside the proton collides head-on with an anti-up or anti-down inside of the antiproton. There are then two key modes for making the Higgs: (1) up + anti-down -> W + Higgs and (2) up + anti-up -> top + anti-top + Higgs. Ultimately, both of these modes are useful only if the machine has sufficient brightness. We are hoping to achieve a brightness in the upcoming Run II (total or integrated luminosity) about 100 times more than what we achieved in the previous Run I which discovered the top quark. [Many of us (including me) think we should push very hard to get the resources to do 1000 times the brightness of Run I. This depends upon $ (dollars) and people]. The purpose is to answer the question: What is the Higgs?

Thus, the Standard Model has taught us that the basic weak scale, the Fermi scale, of mass determines the masses of all other things we see in the Universe. The electron mass, so fundamental to chemistry, biology, the brightness of the sun, etc., derives from the scale of weak force, but through a very feeble coupling to the Higgs condensate. We do not know how this feeble interaction comes about.

At Fermilab in the mid 1990's we discovered the heaviest of the elementary particles, the top quark. As mentioned above, the top quark has a mass very close to 175 GeV, and very near to the scale of Fermi's weak theory. The coupling strength of top to the Higgs is large, it equals 1.00, compared to the lowly electron with it's coupling of about three millionths. Large couplings are often easier to understand because they imply stronger affinities between certain of the various particles in the full theory, and they are easier to connect together experimentally. Therefore, for the first time we feel we may be getting close to an understanding of the origin of the mysterious physics that is responsible for mass itself.

Many theorists are studying the possiblity that the top quark, together with additional new forces, plays a more intimate role in establishing the scale of the weak interactions, and thus the masses of all elementary particles we have measured previously. Key ideas, borrowed from the BCS theory of superconductivity and the known behavior of the strong interactions in particle physics, are being applied in one avenue of investigation known as "top quark condensation" models. Here the Higgs particle is actually a boundstate of the top and anti-top (or of top and anti-X, where X is something new), quarks due to new strong interactions (called topcolor). In recent years these models, which descend from older "Technicolor" models, have overcome some severe difficulties and now are consistent with existing Standard Model data. This is due entirely to facing the challenge of the heavy top quark. The models contain low mass states, called Pseudo-Nambu-Goldstone Bosons, and these may show up in Higgs boson searches first.

Another line of reasoning, and the majority view at present amongst theorists, is "Supersymmetry". Supersymmetry is a hypothetical extension of our understanding of space and time to include additional dimensions that are "fermionic." This means that the dimensions themselves have wierd fermionic properties. For example, a particle that is a boson, such as the photon, when pushed in the direction of a fermionic dimension becomes a fermion, called the photino. Or, a quark, which is a fermion, when pushed in the direction of the fermionic dimension, becomes a boson, called a "squark." So, supersymmetry predicts that for every fundamental observed fermion (boson) in nature there must exist a corresponding "superpartner" boson (fermion). We don't yet see these "superpartners" in nature, so if supersymmetry is a valid symmetry, something must be hiding it at the relatively low energies where we make our observations.

Supersymmetry is intimately connected with theories of quantum gravity, called superstring theories. It can be used to construct supersymmetric theories of the weak scale, but the connection is less mandatory than it is in the connection to gravity.

There are an infinite number of possible supersymmetric models of the weak scale, but one has become a standard: The Minimal Supersymmetric Standard Model (MSSM). The MSSM predicts that all of the superpartners of the quarks, leptons, and gauge bosons (for example, the gluino is the superpartner of the gluon, anf the wino the superpartner of the W, etc.) should be observable, and soon! Both the Tevatron and the LHC will have ample opportunity to begin to see these objects. The MSSM also predicts 5 observable Higgs bosons, but it is expected that only the lightest one will be relatively easy to find. Here, too, the top quark mass has considerable influence on these hypothetical Higgs particles. Without a heavy top quark, the Higgs condensate could not form in the MSSM! The MSSM is fairly specific about where the lightest Higgs must be in mass; it places the lightest Higgs particle in a well defined mass range of less than 140 GeV, well within reach of the Tevatron (and perhaps the reason why LEP was seeing a few funny events at the end of it's run).

In many ways the discovery of the top quark, and the upcoming discovery of the Higgs, marks the beginning of a new era in physics -- the "post-Fermi era" -- in which new discoveries will reach beyond the Standard Model and beyond our current understanding and imagination. The top quark, and certainly the Higgs, will be the key that opens the door to this new era. To this day we have never failed to make a new and unexpected discovery within a new order of magnitude range of energies in subatomic physics. Indeed, the range of energies accessible to the Tevatron may yield the first physics departures from the standard model. The promise of the understanding of something as fundamental to nature as the origin of mass lies ahead. Indeed, we will finally understand the origin of the fundamental constant that Fermi scribbled down some sixty years ago, when steam locomotives and horse-drawn plows were common sights on the American plains, and television and credit cards nonexistent. What novel devices this will lead to in some distant future era will no doubt make the PC-internet era look to them like the horse and steam era looks to us.

High Energy Physics is the study of the forces and the behavior and structure of matter at the shortest distances. It is the ultimate microscopy. The great particle accelerator, the Fermilab Tevatron, is the world's most powerful microscope. Today physicists are looking at the substructure of subatomic particles on a scale that is smaller relative to the atom than the atom is small relative to a human being. In a very real sense, we are now examining, and coming to understand, the very "genetic code" or the "DNA" of matter itself. What could be more fundamental?

-CTH

The Homepage.

I. Introduction.

II. What is Symmetry?

III. Symmetries of Space and Time.

IV. Special Relativity

Emmy Noether

Symmetry in Physics: Chemistry

The Elements

It is generally agreed that the ubiquitous chart, the Periodic Table of the Elements (or the Mendele'ev Chart), is the starting point to the subject of Chemistry. We have discovered that very few high school students and high school teachers understand why the Periodic Table is what it is.

What is the Periodic Table of the Elements? This beautiful HTML version of the Periodic Table of the elements is available at a Los Alamos National Laboratory educational website devoted to chemistry (click here to visit this site). There are many additional educational features about chemistry available there, and you can always get back here by the "back" button on your browser.

The Periodic Table of the Elements

	1 IA 1A	2 IIA 2A	3 IIIB 3B	4 IVB 4B	5 VB 5B	6 VIB 6B	7 VIIB 7B	8 -------	9 VIII 8	10 -------	11 IB 1B	12 IIB 2B	13 IIIA 3A	14 IVA 4A	15 VA 5A	16 VIA 6A	17 VIIA 7A	18 VIIIA 8A
1	1 H 1.008																	2 He 4.003
2	3 Li 6.941	4 Be 9.012											5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18
3	11 Na 22.99	12 Mg 24.31											13 Al 26.98	14 Si 28.09	15 P 30.97	16 S 32.07	17 Cl 35.45	18 Ar 39.95
4	19 K 39.10	20 Ca 40.08	21 Sc 44.96	22 Ti 47.88	23 V 50.94	24 Cr 52.00	25 Mn 54.94	26 Fe 55.85	27 Co 58.47	28 Ni 58.69	29 Cu 63.55	30 Zn 65.39	31 Ga 69.72	32 Ge 72.59	33 As 74.92	34 Se 78.96	35 Br 79.90	36 Kr 83.80
5	37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc (98)	44 Ru 101.1	45 Rh 102.9	46 Pd 106.4	47 Ag 107.9	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.3
6	55 Cs 132.9	56 Ba 137.3	57 La* 138.9	72 Hf 178.5	73 Ta 180.9	74 W 183.9	75 Re 186.2	76 Os 190.2	77 Ir 190.2	78 Pt 195.1	79 Au 197.0	80 Hg 200.5	81 Tl 204.4	82 Pb 207.2	83 Bi 209.0	84 Po (210)	85 At (210)	86 Rn (222)
7	87 Fr (223)	88 Ra (226)	89 Ac~ (227)	104 Rf (257)	105 Db (260)	106 Sg (263)	107 Bh (262)	108 Hs (265)	109 Mt (266)	110 --- ()	111 --- ()	112 --- ()		114 --- ()		116 --- ()		118 --- ()

Lanthanide Series*	58 Ce 140.1	59 Pr 140.9	60 Nd 144.2	61 Pm (147)	62 Sm 150.4	63 Eu 152.0	64 Gd 157.3	65 Tb 158.9	66 Dy 162.5	67 Ho 164.9	68 Er 167.3	69 Tm 168.9	70 Yb 173.0	71 Lu 175.0
Actinide Series~	90 Th 232.0	91 Pa (231)	92 U (238)	93 Np (237)	94 Pu (242)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (249)	99 Es (254)	100 Fm (253)	101 Md (256)	102 No (254)	103 Lr (257)

** Groups are noted by 3 notation conventions.

For a list of a the element names and symbols in alphabetical order, click here

This unit is an "arm waving" effort to explain, from fundamentals, how the Periodic Table gets to be the way it is, and, in so doing, touches on some quantum theory and the all important, but also all too unreasonably ignored, concepts of symmetry.

We start with something that should be taught in low level courses in Physics and Chemistry. It may take some major swallowing on the part of the student. But it will be rewarded.

The Atom of Niels Bohr

The Hydrogen atom was first conceived of in quantum theory by Niels Bohr. Now, this is a chicken and egg situation. Which came first, the Quantum Theory or the theory of the Hydrogen Atom? In essence they both came along together. The theory of the Hydrogen atom was a major conceptual leap forward for the infant quantum theory, which no one had a clue as to how to construct. It was also a major step forward for the understanding of the atom itself, and hence the science of Chemistry, which had previously defied understanding within the classical theory of physics.

Of course, some parts of the atom were known, more or less, from empirical studies. It was known from Rutherford that there is a tiny core inside of an atom, called the nucleus, wherein resides 99.98% of the mass. And it was more-or-less realized that electrons, discovered in 1898 by J. J. Thompson, somehow coexist in space with the nucleus. And, it was clear that electromagnetism somehow holds this system together.

But, according to Maxwell's theory of electromagnetism, when placed in a classical circular orbit, the electrons immediately radiate all of their orbital energy away and collapse down onto the nucleus. Such an object would be chemically dead. In fact, such an object makes no sense. In fact, nothing about electrons, atoms or nuclei makes sense at all in the classical theory. So, we need the quantum theory. One nice simplification happens here, however. We can neglect the effects of Relativity because the electrons are actually moving at velocities that are slow compared to the speed of light as they move about the nucleus (the move at a few one thousandths of the speed of light in their motion). However, the motion is nothing whatsoever like that of a planet going around a star. It isn't possible to describe the motion as a precise orbit, where we say: "the electron is at position x traveling with velocity v at time t." We must settle for something containing less information than a classical orbit; we must settle for the quantum orbital, or the wave-function, which determines the probability of finding the electron at x and time t.

We'll assume that a preliminary course in conceptual physics or chemistry gives students a sense of the structure of atoms: atoms are a central, very small nucleus, carrying almost the entire mass and the total positive charge of the atom. Outside the nucleus are the negatively charged electrons in cloud-like orbits about the nucleus. The laws of quantum theory teach us --- and we transmit this to the chemistry students --- that only certain special orbits of the electron motion, or orbitals, are allowed. The orbits are "quantized," and hence the term "quantum theory." From Louis de Broglie's 1923 insight, we learn that electrons behave like waves, and each electron is associated with a "wavelength." Experiments carried out by G. P. Thompson in England in 1928, and by C. Davisson and L. H. Germer at Bell Laboratories in New Jersey confirmed DeBroglie's theory that the wavelength "associated" with an electron is given by:

$\lambda = h/mv$

Here h is Planck's constant, the "logo" of quantum theory, and m is the electron mass and v the electron velocity. Indeed, mv is just the electron's "momentum."

In the old quantum theory of Bohr we assume that when an electron is bound in an atom it moves in a circular orbit of radius R . The allowed orbits are only those in which the circumference of the orbit, $2\pi R$, matches to an exact integer number of wavelengths. This is much like the sound vibrations produced by a musical instrument, e.g., the "standing waves" that occur on a guitar string or in an organ pipe have wavelengths that are (inverse integer) multiples of the length of the string or pipe. Putting the statement mathematically, we can squeeze n wavelengths of the electron into one full circumference, or:

$n\lambda = 2\pi R = nh/mv$

where n is an integer: 1 , 2 , 3 , etc., which enumerates orbits and λ is the wavelength of the electron, given by eq.(1.1).

We can now derive a formula for the allowed radii of the orbits in terms of other known quantities. Our delicious mixture of quantum and classical ideas leads to an expression for the orbit radius " R " in the atomic orbit that depends only upon n (i.e., we eliminate the unknowns, λ , v , etc.).

We start with Newton's law of motion F=ma , and we assume a circular orbit. For a particle moving in a circular orbit of radius R with velocity v there is a "centripedal" (pointing toward the center) acceleration of v^2/R . Thus, to produce the motion in the circular orbit we need to apply a force of $F = ma = mv^2/R$ directed toward the center of the orbit. Our centripedal force is provided by the electrostatic attraction, the "Coulomb force," between the negatively charged electron and the positively charged nucleus of the atom. This is given by Ze^2/R^2 , for the force between the nucleus of charge Ze and the electron of charge e , separated by a distance R . (Note: Here, the charges are measured in the centimeter-gram-second system of units in terms of "esu" or "electrostatic units;" The electron has a charge of 4.8×10^{-10} esu)

So, we have:

$mv^2/R = Ze^2/R^2$

but eq.(2) tells us that v is related to R :

$v = nh/ 2\pi m R$

We substitute the expression for v in eq.(4) into eq.(3) and solve for R:

$R = n^2h^2/4\pi^2 m Z e^2$

Voila! This is the desired expression for the atomic orbit radius in terms of known quantities and the "quantum number" n , where n = 1, 2, 3, ... For the special case n=1 we define the expression of eq. (1.5) to be the symbol a_0 :

$a_0 = h^2/4\pi^2 m Z e^2$

The quantity a_0 has dimensions of length, and is called the "Bohr radius." When we put in Planck's constant, the π's, etc., we get:

$a_0 = 5.25 \times 10^{-9}$ cm

So we see how we get only certain allowed radii for the orbits, given by:

$R = n^2 a_0$

This is completely different than in the case in classical physics, where we could have any radius or any orbit we desire by simply choosing the velocity of the electron at will. We say that the orbits are "quantized." For the special case of the Hydrogen atom we have Z = 1 , and the orbital radii are then: a_0 , $4a_0$, $9a_0$ ...

What about the energy of the orbiting electron? The total energy of the electron is the sum of the kinetic energy, the "energy of motion," $m v^2/2$, and the potential energy, $-Ze^2/R$. Note that the potential energy is negative here because the force is attractive between the nucleus and electron; thus the electron is pulled down into the negative potential; a positive potential energy would represent a repulsive force, such as between pairs of electrons. Hence, for the total energy we have:

$E = m v^2/2 - Ze^2/R$

However, from equation (1.3), we see that $mv^2/2 = Ze^2/2R$, i.e., the kinetic energy equals half the (positive) magnitude of the potential energy. So, upon substituting into eq.(1.9), the formula for the total energy, we have:

$E = -Ze^2/2R$

Now, using our formula for R in eq.(5) we get:

$E = - 2\pi^2 Z^2 e$^4/$n^2 h^2$ = - 13.6 electron volts x Z^2/n^2

Thus, each allowed orbit has an energy that is associated with with the quantum number n . For the special case of the Hydrogen atom, we have Z=1 and therefore the energies are:

(n = 1) $E_1 = -13.6$ eV

(n = 2) $E_2 = -3.4$ eV

(n = 3) $E_3 = -1.52$ eV

What is the meaning of negative energy? This means that the electron is bound to the nucleus in an atom. Thus, if we want to free the electron from the "groundstate" (the state of lowest energy, or most negative energy), we must give it at least 13.6 eV of energy. Then it can escape. So -13.6 eV is the lowest possible energy of an electron bound in a Hydrogen atom, and therefore the most stable state.

We learn that once it is in an excited state, an electron can jump, or fall, down to a lower energy orbit, releasing a quantum of light energy: a "photon." Conversely, an electron can jump from a lower energy to a higher energy orbit, if it absorbs the required energy from some outside source of photons. The quantum theory predicts the exact values of these photon energies to be the exact differences between the energies of the levels of the hopping electrons. Thus, Bohr's simple theory explains the discrete spectral lines of ionized gases, such as in a discharge tube in the lab, or in the corona of the sun. This had been a complete mystery for the preceding half century of physics.

We apologize that here we used the "old quantum theory" of N. Bohr, rather than the modern (and complete) version based upon the Schroedinger Equation, with the electron described by its wave function $\psi(x, t)$. We all know that the Bohr theory we have outlined above gives the right answer and is a quick study that every student (and teacher) should know. In the full quantum theory, however, the electron does not circle the nucleus in a well defined orbit, but rather is described by a wave function, $\psi_{n,...}(x, t)$. The wave-function determines the probability of finding the electron in the n th orbital, at the location, x , at any time, t , given by: $|\psi_{n,...}(x, t)|^2$. All statements in the old quantum theory are essentially sort-of "averages" as defined by the more complete theory. We'll have more to say about this below.

More Quantum Numbers

As we have seen, the quantum theory typically selects only certain discrete energies or orbits as contrasted with Newtonian (classical) theory where arbitrary energies and orbits would be allowed. Discrete values of the orbit are enumerated in the quantum theory by integers (or half-integers) called quantum numbers. Thus in eq.(1.5), n is a quantum number which can take on integer values 1, 2, 3, . . . for the different radii. n is special, and is called the "principal quantum number."

It turns out, however, that other properties of the atom, which would have classically a continuum of values, are also "quantized" and are allowed only certain values. One important property is angular momentum which in Newton's mechanics is a vector quantity given by: L = (r) x (p). The magnitude of this vector is given by |L| = mvR for the case of an electron in a circular orbit in classical physics. In Quantum Mechanics, the magnitude of L , too, is quantized with a new quantum number l, and the formula for |L| is somewhat more complicated:

$|L| = (h/2\pi)\ (l(l + 1))^{1/2}$ (l = an integer)

where we introduce the more conventional quantity, \hbar = $h/2\pi$. To understand the orbital angular momentum in the atom, and the above formula, one really has to work through the full solution of the Schroedinger equation, which is beyond the scope of this article (we refer you to the standard book by L. Schiff [2] or any other reasonable book on Quantum Mechanics; it requires a basic understanding of ordinary and partial differential equations, and their special function solutions). We'll just state the results presently. The angular momentum quantum number, l , in an orbital of principle quantum number n , can take on values from zero to a maximum value which is n-1 . one less than the principal quantum number. For example, if n = 1 , then l = 0 . If n = 2 , the l can take on values of 0 and 1 ; if n = 3 then l can take on values of 0 , 1 and 2 , and so on.

Angular momentum is further complicated by the fact that, because it is a vector, it can therefore can point in some direction in space. In Quantum Mechanics the question of measuring the direction of a vector like angular momentum is subtle and is posed somewhat philosophically, i.e., in any experiment the experimentalist chooses a direction in space, e.g., the "z" direction, and asks: What is the value of the angular momentum pointing in this direction? Again, the answer is always quantized: For any value of l , the measured value of angular momentum along, e.g., the z-direction is

L_z = m \hbar where (m = -l, -l+1 -l+2 , . . ., +l - 1, +l)

For example, an orbit can have angular momentum l=1 , hence the magnitude of the total angular momentum is |L| = \hbar $(1(1+1))^{1/2}$ =\hbar$(2)^{1/2}$. Then the experimentalist picks a direction, e.g., the y-axis, and she will measure a value of either \hbar , 0 or -\hbar along this axis, corresponding to the three allowed values of m, of (1,0,-1). Only when l becomes very large do we approach the classical limit. For example, with l = 1000 , we can have a measured value (a "projection") in the z-direction of m=5 (this is like having a classical angular momentum vector nearly perpendicular to the z-direction); alternatively we can have l=1000 m=-998 (this is like having the classical vector nearly pointing exactly in the -z direction), etc.

Finally, of quintessential importance, the electron itself has internal angular momentum, called spin. Spin, which sounds like what a top does, is really spooky; it is an intrinsic angular momentum which is simply a part of the electron. We can never stop an electron from spinning. We know the value of the electron spin angular momentum has l_e = 1/2 , which is always a fixed value, and never changes. Then, the measured value of electron spin along the z-direction can be only one of two values (which we denote by s , the analogue of m: either s = +1/2 (often called "spin up"), or s=-1/2 (often called "spin down"). If all this is somewhat unsettling to you, join the quantum crowd. We physicists become comfortable (eventually) using Quantum Mechanics, because it is true and it works, but we never really become comfortable that we fully understand it.

Thus, in the quantum world of an electron in an atom, there are four quantum numbers:

n, l , m, s

Let's hold the many questions and accept that these four quantum numbers completely define the state of motion of an electron in an atom. From these quantum numbers and the mathematical solution of the quantum equations, we get a complete description of where the electron is in its various states.

As we mentioned above, in the full quantum theory, we can only compute the wave-function, ψ , which, when squared, gives the probability of finding an electron in some small volume somewhere at some time in the space of the atom. ψ is a complicated function of position, (x,y,z) and time, t . This is written as ψ (x,y,z,t) . Each orbital is specified by n,l,m,s , so we write the wave-function for any given orbital as $\psi_{n,l,m,s}$(x,y,z,t) . In general, in quantum mechanics, the description of any physical system e.g. a quark, atom, molecule, two particles undergoing a collision, etc., are described by some kind of "wave function" in analogy to ψ. This function will in general depend on all the coordinates of all the particles in the system and will contain whatever other relevant properties exist of the particles, e.g., charges, spins, etc. To obtain predictions as to the results of measurements, we must take the absolute value squared of ψ which gives the probability of obtaining results from the measurements.

This is half of what we need to know to construct the Periodic Table. There's more, and it revolves around the fact that all electrons are exactly identical in all respects. Let's proceed.

Enter Symmetry

Now that we have given a very limited view of how quantum theory works, we want to use the concept of symmetry to understand how nature builds up the chemical elements, atoms, from the simplest: Hydrogen with one nuclear charge (Z = 1) and one electron, all the way to Uranium with (Z = 92) 92 electrons and beyond.

We have noted that there are states of discrete energy, the lowest energy (n = 1) being the most stable. For an atom of nuclear charge Z , e.g., Z = 2, 3, 4 . . . (i.e. Helium, Lithium, Boron, ...) we must add electrons in order to balance the nuclear charge. Like a marble on a sloping floor which rolls down to the potential energy minimum, the electron will roll down to the lowest energy state of allowed motion --- so too will the next 10 marbles. Where do they go?

We now have to understand the rules. The rules emerge from the concept of symmetry. We define symmetry as follows: a system exhibits symmetry when it does not change even though you perform some operation on it. A perfect cylinder can be rotated through any angle around the symmetry axis leaving the system identical to how it was before the rotation. A sphere exhibits an even more perfect symmetry; it doesn't change if you rotate it about any diameter through any angle. We call this rotational symmetry. A 3-bladed propeller, or an equilateral triangle, has a more restricted symmetry; they doen't change if you rotate the through 120^o about the center symmetry axis.

The fancy language is that the system is "invariant" (doesn't change) when subject to some symmetry operation. Any physical system is described by certain mathematical equations, e.g., we have $x^2 + y^2 + z^2 = R^2$ for the points defining a sphere of radius R in an (x,y,z) coordinate system. We can perform a rotation on the coordinate system about the origin. After the operation of rotation, each of the coordinates change x --> x' , y --> y' , and z --> z', but the symmetry of a sphere implies that after the rotation we still get: $x'^2 + y'^2 + z'^2 = R^2$. The mathematical description of the sphere hasn't changed, or is "invariant" under the rotation.

The symmetry we want to discuss presently has to do with two electrons, one located at x1 and the other located at x2. There is a hypothesis that all electrons are identical in every detail. Thus, if we interchange the locations of the two electrons in space, x1 <---> x2 the system looks the same, indeed, it is the same! This is a symmetry: it is called "exchange symmetry."

How do we describe exchange symmetry mathematically? The description of our two electrons is given by a wave function ψ(x1,x2, t) (we'll henceforth not write t). If we interchange the two electrons, we interchange x1 and x2 we will get the new wave-function ψ(x2,x1) . But, The result of this interchange, according to our hypothesis, must be an identical system! So, there must be no change in the probability of observing the two electrons, one at x1 and one at x2 , i.e.,

$|\psi(x1,x2)|^2 = |\psi(x2,x1)|^2$

At the wave-function level there are two ways of satisfying this equation:

either (A): $\psi(x1,x2) = \psi(x2,x1)$

or (B): $\psi(x1,x2) = -\psi(x2,x1)$

(this is just the fact that $4^{1/2}$ has two solutions, +2 and -2). Solution (A) says that ψ is an even function under the swapping of the positions of the two particles, sometimes called a symmetrical function. Solution (B) says that ψ is an odd function under swapping positions, called anti-symmetric. So what?

It turns out that nature makes a really big fuss about even vs. odd behaviors under exchange of identical particles. A broad class of particles, including electrons, protons, neutrino's and quarks, obey Solution (B), i.e., they are anti-symmetric under interchange of any two particles. Yet, another class of particles, photons, pions, W's, Z's, . . ., behave as in Solution (A) and are even functions under exchange. There is a very deep connection: the first list all have a spin equal to 1/2, 3/2, 5/2, . . . i.e. odd half integer spins. The second class have spins 0, 1, 2, 3 . . . integral spins. These two groups behave very differently. The half integer particles are collectively called "fermions," the integer spin objects are called "bosons."

Now here is how they behave differently. The odd particles have the property that if all of their quantum numbers are the same (charges, spins, etc.) and we attempt to set x1 = x2, i.e. the two particles are pushed to the same point in space at the same time, then (4) tells us that

$$\psi(x1,x1) = -\psi(x1,x1)$$

which can only be satisfied by $\psi(x1,x1) = 0$. Therefore, the probability of identical electrons co-existing at the same place in space at the same time, is zero (their spins must be aligned in this case, i.e., s1 =s2)! It is as if some powerful force prevents them from co-existing at the same point in space. Lest you say: "of course, it is just the electrical repulsion of negative charges," we remind you that the prohibition against co-existing in the same point in space simultaneously also applies to neutrons and neutrinos!

The particles of the boson group, on the other hand, have no problem co-existing at the same point in space (again, for identical spin states, charges, etc.) because $\psi(x1, x2)$ doesn't have to vanish at x = y for the even symmetry. In fact, when many many particle states of bosons are created there is an enhanced probability of all of the bosons going into exactly the same state of position, or momentum. You may have read about Einstein-Bose condensates, i.e. clusters of boson particles that form a very compact and dense coherent state; laser beams are coherent states of many photons in the same exact state of motion, while superfluids are coherent states of bosonic He^4 marching together in bosonic lock-step.

The apparent attraction of bosons to each other and the apparent repulsion of fermion particles give rise to the descriptive term "exchange force," even though there really is no force acting here, only the curious behavior of the wave functions respecting the symmetry of identical particles. The exchange force, which is not a real force, requires some justification. The wave function and its square, the probability of a given arrangement of two electrons, gives us some feeling for this. If we consider a situation which has a high probability, e.g. throwing a seven in a pair of dice, the behavior is as if the "3" and the "4" have an attraction for each other, or, in general, the dice are much more "attracted" to the "7". Similarly, they are least attracted or even repelled from the "2" and the "12", the least probable numbers.

So very probable outcomes i.e., $|\psi(x,x)|^2$ close to unity, appears to act like an attractive force, whereas $|\psi(x,x)|^2$ close to zero appears as a repulsive force. Physicists call this the exchange force (which is not a force!).

Wolfgang Pauli & the Elements

Now comes the question of building up the elements in the Periodic Table. We can discuss this by

considering a nucleus of Z = 2 (Helium). We need to add a second electron to the atom whose nucleus has charge +2. One would naturally assume that the electron would go to the lowest energy orbit, n = 1. However, we know that two negatively charged electrons repel each other and thus, it may come out that the second electron is happier (the total energy is lower) if it goes into the n = 2 orbit. After some analysis, however, it turns out that there would have been plenty of room for more electrons to pile into the the n = 1 orbit, as Z is increased, minimizing the total energy. Thus Helium will have two electrons in the n=1 orbits. As we increase Z , then, are all atoms just fatter and fatter Hydrogen atoms, with all electrons in the n=1 orbit? For example, does the Iron atom with Z=56 have 56 electrons squashed into the n=1 orbit?

The answer is no! We have just seen that here is an "influence" in nature, much stronger than the Coulomb force, which determines where the second, and third, and fourth, ... electrons cannot go. It is just the odd property of the wave-function under exchange of two electrons, and it implies the dramatic result:

No two electrons can co-exist in the same identical quantum state.

This "influence" is called: The Pauli Exclusion Principle, after Wolfgang Pauli, the Austrian genius who did most of his research at the Swiss University, ETH, in Zurich, and helped to build the modern quantum theory in the early 20th Century. This crucial principle determines why The Periodic Table is what it is; it also determines the future of stars and planets and protons and people. Pauli's discovery ultimately gives rise to the strict set of rules as to how we go from the simplest element (Hydrogen of Z = 1), to the complex elements with Z = 2, 3, ... 92, ... 110 ... . In German, it is the word "Aufbauprinzip"; in English translation, it is the "Principle of Building Up (the Elements)".

In the full quantum theory description of an electron in an atom, the electron is completely described by its four quantum numbers: n, l, m, s . The Pauli Principle then says: "no two electrons in an atom can have the same four quantum numbers." *This explains the Periodic Table of Elements*. It prevents all the electrons from going to the lowest energy state, as they would in a world ruled by classical physics. It also influences the chemical activity of an atom by determining whether or not electrons can be shared between atoms. Thus, the columns of the Periodic Table of elements, which represent the common chemical properties of the atoms within a given column, is controlled by the exclusion principle.

As the simplest example, the electron in atomic Hydrogen goes to the n = 1, l= 0, m = 0, s = 1/2, state. In Helium we need to add another electron. It too can go into the n = 1 state (which forces l=0, m = 0) only if its spin is s= -1/2 , i.e. opposite to the first electron. This exhausts all the possibilities of the n = 1 state. We thus have a "closed shell." Lithium has three electrons, the first two huddle in the n = 1 state, i.e., the closed shell structure of Helium. The third electron must go into the n = 2 state.

Beyond Lithium, how many electrons can go into the n = 2 state? First we fill the l = 0 subshell with two electrons, one with s =+1/2 , one with s =-1/2 . This shell has m = 0 Then we fill the l = 1 , m = -1 state with 2 electrons. Next, we can put 2 electrons in the l = 1 , m = 0 , state and finally 2 electrons in the l = 1 , m = -1 state. The total number of electrons allowed into the next n=2 shell is 8. This takes us from Lithium through Beryllium, Boron ... to Neon with Z = 10 . All of this is illustrated in table I.

Not only do we build up all the elements this way, but we get to understand the chemical properties of the elements. Hydrogen has room for one electron in the n = 1 state. It therefore forms compounds with elements that have a spare electron it can share in the n=1 shell, e.g., H_2 or Li H , Lithium Hydride. Once a shell is completely filled, as in n=1 Helium, or n=2 Neon, or n=3 Argon, etc., we have an inert or noble gas.

To follow the "aufbauprinzip" in detail requires a knowledge of the rules which connect the various quantum numbers. These emerge from the solution of the quantum theory equivalent of Newton's Laws, the Schrodinger equation. The rules are not important for one to grasp the central idea, electrons will try to go to the lowest energy state, but they must be consistent with the Pauli rule.

Although we noted that the various n-states differ from each other in energy, there are smaller energy differences between different l-values and even m-values. The problems get pretty complex when we are deciding how electrons are shared between atoms in a molecule. Larger molecules require computer solutions and quantum computational chemistry is a hot subject these days.

The exchange forces (which are not real forces) play a major role in the theory of molecular bonds --- the heart of chemistry. The Pauli Exclusion Principle also shows up in the detailed construction of atomic nuclei, a crowded little volume of space filled with spin-1/2 neutrons and protons. It shows up in such astronomical processes which govern the life cycle of stars, from supernova, neutron stars to black holes. You did good, Professor Pauli!

It is rather remarkable that the identity of particles gives rise to the 100 or so chemical elements, which in turn give rise to billions of possible molecules and gives our world variety and richness!

As we have seen, the patterns of atoms, the properties of chemistry, indeed much of the stability of matter itself, is governed by the Pauli principle. Where does it come from? We have only asserted, but not proved, that odd half-integer spins, 1/2, 3/2, 5/2, ... , (known as "fermions") obey the odd solution under exchange of position, of eq(17), while the integer spins, 0,1,2,... (known as "bosons") go with the even solution.

Indeed, the Pauli exclusion principle actually follows from deeper symmetries of physics, in particular, the symmetry of "rotational invariance." Sometimes the Pauli principle is called the "spin and statistics theorem."

Spin and Statistics

We will sketch the idea here.

As we saw for the electron, all elementary particles have spin, and are described by wave-functions that depend upon position in space. The wavefunctions generally change under rotations of the coordinate system. For simplicity, let us assume that two electrons occupy the same orbital, and the position dependence of the wave-function of the orbital is spherically symmetric about the origin, (this means l =0 , hence m=0) so rotating the spatial location of the particle has no effect on the wave-function. There remains, however, an effect from the spin of the particle.

The change of a spherically symmetric wavefunction of a single particle under rotation is given by the spin quantum number s (recall that s is the spin projection in the z -direction of space).

The wave-function, ψ_s(x1), changes as we make a rotation through an angle θ about the z axis by an amount: $\psi_s(x1) \rightarrow e^{i\theta s}\psi_s(x1')$ where $i=(-1)^{1/2}$. Suppose our particle is a boson with s=1 and we rotate the wavefunction through 360^o or 2π radians. Then: $\psi_s \rightarrow e^{i\theta s}\psi_s = e^{2\pi i}\psi_s = \psi_s$ That is, since s=1 , then $e^{2\pi s i} = e^{2\pi i} = 1$. This is not surprising; when we rotate any classical object through 2π we expect it to come back to its original position.

What happens to an electron? An electron is described by a weird mathematical object called a "spinor." A spinor can be thought of as the "square-root of a vector." Consider an electron which also describes spin pointing along the z -direction. Let us rotate its spinor through 2π about this axis. The spinor has s=1/2 , and we find:

$\psi_{1/2} \rightarrow e^{2\pi s i}\psi_{1/2} = e^{\pi i}\psi_{1/2} = -\psi_{1/2}$

The spinor changes into minus itself when we rotate through 2π ! This is seemingly impossible! We have to rotate the spinor though 4π to bring it back to itself. Thus, intrinsically through the sign of the wave-function every electron knows if it has been rotated through one cycle of 2π . Is this effect observable? Certainly, the probability of finding the rotated electron is the same as the unrotated electron, since the probability is just the square of the wave-function and $(-1)^2 = +1$. So, the behavior of the spinor is really happening at the quantum level, and goes beyond classical physics! Yet, this effect is observable and underlies the Pauli exclusion principle, and is therefore profoundly observable!

Now, suppose we consider two identical electrons with equal spin values of s1 = s2 = s , one located at x1 = (x,0,0) and the other at x2 = (-x, 0, 0) , both is the same spherically symmetrical orbital at the same time. This state is described by a wave-function $\psi(x1, x2)_{s,s}$. Since the electron's spins are identical, it must be possible to exchange the electron's positions and get back the same wave-function with the positions interchanged without exchanging the values of spins:

$\psi(x1, x2)_{s,s} \rightarrow \psi(y,x)_{s,s}$

But, if we think about this interchange for a minute, we see that we can get the same result by rotating the system by 180^0 or π radians about the z axis, which interchanges x and y . This wave-function then changes by

$\psi(x1, x2)_{s,s} \rightarrow e^{\pi s1\, i}e^{\pi s2\, i}\psi(y,x)_{s,s} = e^{2i\pi(s)}\psi(y,x)_{s,s} = -\psi(y,x)_{s,s}$

where each factor of $e^{\pi s i}$ takes care of the rotation effect on each spinor. So, we find that under the interchange of the electrons, the wave-function has changed sign:

$\psi(x1, x2)_{s,s} = -\psi(y,x)_{s,s}$

Our electrons with half-integer spins are therefore fermions! And, if we try to set x=y we must get $\psi = 0$.

If our particles had integer or zero spin we would get:

$\psi(x1, x2)_{s,s} \rightarrow e^{\pi s i}e^{\pi s i}\psi(y,x)_{s,s} = e^{i2s\pi}\psi(y,x)_{s,s} = +\psi(y,x)_{s,s}$

This is the essential statement of Pauli's exclusion principle.

This kind of argument only works in certain special configurations, e.g., above we specialized to two identical s_i , and spherical orbitals. It can be generalized, however: Since the spin and interchange symmetry must work in the special case of s_1=s_2=s , then it must also be true for the general case of arbitrary s_1 and s_2 , because we can rotate one electron relative to another if only weak forces act between them, and this should not disrupt the symmetry of the overall wave-function. The relative exchange symmetry must also hold for any number of particles making up the overall state because, in the absence of strong forces, we can always consider any pair of electrons in the system to be in approximate isolation from all the others. The only unsettling feature of this "proof" is that really strong forces might undermine it. Indeed, people used to fret about whether or not quarks obeyed the normal Pauli exclusion principle. Even the strong forces, however, that bind quarks are relatively weak at short distances, so quarks too obey the Pauli Exclusion Principle.

Finally, you might find it unsettling that we disallow a state which becomes minus itself under an exchange (e.g., two electrons with the same spins at the same point in space), while do allow single electrons, which are spinors and which become minus themselves under rotation through 2π ! One can, however, get very philosophical about this, argue that single electrons don't really exist, invoke Mach's principle and rotate the Universe around the electron, etc., etc. It is best, however, to simply accept the Pauli principle at face value: it is a fact about nature that is verified by billions and billions of atoms and many thousands of experiments.

[1] See for example, D. Ebbing, *General Chemistry*, (Houghton Mifflin Co., Boston 1984)

[2] L. Schiff, *Quantum Mechanics* , (McGraw-Hill, 1968).

[3] R. P. Feynman, The Feynman Lectures on Physics, (Addison-Wesley Pub. Co., 1963).

[4] For limits on time dependence of fundamental constants see, e.g., F.W. Dyson, in: *Aspects of Quantum Theory* , eds. A. Salam and E.P. Wigner (Cambridge Univ. Press, Cambridge, 1972) 213; in: *Current Trends in the the Theory of Fields* eds. J.E. Lannutti and P.K. Williams (American Institute of Physics, New York, 1978) 163. See also, C. T. Hill, P. J. Steinhardt, M. S. Turner Phys.Lett., B252,1990, 343, and references therein.

[5] see, e.g. *Women in Mathematics* , L.M.Osen, MIT Press (1974) 141.

Symmetry in Physics: Portraits of Emmy Noether

There are many interesting websites that feature biographical information about Emmy Noether. Here are some of our favorites:

Emmy Amalie Noether and other famous mathematicians and scientists are decribed in these biographical sketches provided by The School of Mathematics and Statistics of the University of St. Andrews, Scotland. This site includes a very nice photo album of Emmy Noether throughout her life.

The Emmy Noether Lectures feature distinguished women mathematicians and are presented by The Association for Women in Mathematics. A biography of Emmy Noether can be found from this page.

See the associated link at the homepage of Sunsook Noh . This is a graphically beautiful site, featuring photographs of past lecturers.

The best of Math News at Waterloo College features an article about Emmy Noether by undergraduate Marni Mishna.

A must see is the Emmy Noether Society of the Math Club of Bowdoin College.

Prof. of Mathematics, Clark Kimberling, of Evansville College provides a group photograph of Emmy Noether, Mentors and Colleagues and a brief biographical synopsis.

There are, of course, numerous other sites worth seeing which can be found in any search engine under "Emmy Noether".

We also recommend the monograph: "Women in Mathematics", by L. M. Osen, MIT Press (1974) 141.

The authors for this website emmynoether.com (Teaching Symmetry in the Introductory Physics Curriculum) are

Prof. Christopher T. Hill,
Theoretical Physics Department,
Fermi National Accelerator Laboratory,
MS. 106, P.O. Box 500, Batavia, Illinois, 60510, USA ***email: hill@fnal.gov*** .

Prof. Leon M. Lederman,
Illinois Math and Science Academy
Resident Scholar
1500 W. Sullivan Rd.
Aurora, Ill. 60506-1000 ***email: lederman@fnal.gov*** .

Emmy Noether

Symmetry in Physics: Proving Noether's Theorm

I remember sitting in the seminar room at Caltech in the early 1970's, waiting for the speaker to arrive. Richard Feynman was there, and an assembly of students and faculty, and an impatient silence permeated the room. A graduate student, an experimentalist I believe, broke the silence and, showing no temerity, asked Feynman: "How do you prove that your Path Integral formulation of Quantum Mechanics is correct?" Feynman replied: "You don't have to `prove it' because it is true! Nature doesn't know about `proof'!"

This shows how physicists differ philosophically from mathematicians. Nature is the ultimate arbiter; the Path Integral works, and Quod Erat Demonstratum! A "proof" is a refinement of the statements we make about nature for the human linguo-conscious logical system. This is not to disavow scientific reasoning, or western scientific philosophy, but just a quip that hand waving arguments are often good physics, and you shouldn't get too hung up on the precise mathematical foundations (unless, of course, you want to be a mathematician). Of course, Feynman did then elaborate about the physical content of the Path Integral, and it was fun and illuminating, and I don't even remember the seminar or the seminar speaker we had that day.

Now we turn to the question of "proving" Emmy Noether's theorem. We won't give any kind of "official proof," but rather a kind of proof by example.... or .... as we often say in physics .... a hand-waving proof. We apologize that the following discussion requires some basic understanding of

Calculus. Calculus was invented by Newton and Leibnitz to describe physics. After a while, the only way to describe the French experience is to learn French. Likewise, after a while, the only way to describe nature is to learn a little Calculus. [It would, we think , be better if the subject of Calculus was taught side-by-side with physics in the High School curriculum. Separating them does a disservice to both subjects. (IMHO; actually this is CTH talking; I'm not sure what Leon thinks about this issue)]

Consider a free particle of mass m. The energy of the particle in Newtonian mechanics is just given by:

$E = m v^2/2 = m(dx/dt)^2/2$

We might ask if there is enough information in the energy formula to determine the equation of motion, where the equation of motion is:

$0 = m(d^2x/dt^2) = m (dv/dt)$

where v is the velocity, $v = dx/dt$. Essentially we are asking: "what is the relationship of the equation of motion to the energy formula?" Of course, we can derive the energy formula from the equation of motion easily enough. We just multiply the EOM by the velocity,

$0 = mv(dv/dt) = m (d(v^2)/dt)/2$

and we integrate this to prove that $mv^2/2$, the energy, is a constant of the motion.

However, mathematicians and physicists like to see if arguments can be run in reverse. Starting with the energy formula, can we reverse engineer the equation of motion? This is a matter of principle. What is truly more fundamental, the energy, or Newton's equation? This requires setting up a slightly different way of thinking about nature. The new way is called the "Action Principle." In fact, it is the best way to think about how nature works.

We begin by introducting the idea of a trajectory, or a path. We say that the particle moves on a trajectory, x(t), starting at x_0 at time t_0, and ending at x_1 at time t_1, that is:

$x_0 = x(t_0)$ and $x_1 = x(t_1)$.

Now, the trajectory is completely arbitrary. It can zig-zag, twist in a helix, roll, or do anything that we can draw or imagine.

We now introduce the concept of the action. The action is defined, *for a free particle* and for any trajectory, as the time integral of the energy (this definition change below when we include interactions):

$$\int dt\, (1/2)m(dx(t)/dt)^2$$

where the integral runs from t_0 to t_1, and our trajectory, x(t), runs from x_0 to x_1. We emphasize that you can compute the action *for any path* that satisfies the intial and final conditions. You should try this for a few sample paths, and play around with it. You are not restricted to the special path that describes the true motion from x_0 to x_1, the one that satisfies Newton's equation (i.e., uniform motion). The path can have wiggles, stops and starts, and so forth. The Action Principle selects the true path for the motion.

The Action Principle: The free particle will move along the particular trajectory from x_0 to x_1 that has the minimum value of the action.

First, we emphasize "free particle " here. The action principle has to be stated more carefully when we include forces and things, and we replace energy by something called the Lagrangian (see below). Moreover, in general, we say the motion "extremalizes" the action, i.e., it could also be a maximum of the action. However, a minimum is an extremal value, and happens to be the correct answer for a single free particle. Ok, but where does Newton's equation come from?

Suppose that we had found a particular trajectory x(t) that has the smallest value of the action. Let us slightly change the trajectory by a tiny amount. Let us add an arbitrarily small change z(t), so the trajectory is now x(t) + z(t). Furthermore, z(t) is so small that we will always neglect $(z(t))^2$, which is even smaller. If we substitute the new trajectory into the formula for the action we get the original result plus a small correction:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 + m(dx(t)/dt)(dz(t)/dt)]$$

The expression can then be rewritten:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 + m\ d(z(t)(dx(t)/dt))/dt - z(t)\ m\ d^2x(t)/dt^2\ ]$$

hence we get:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 - z(t)\ m\ d^2x(t)/dt^2\ ] + mv(t_1)\ z(t_1) - mv(t_0)z(t_0)$$

The new path also starts at x_0 and ends at x_1 so we need to fix $z(t_0) = z(t_1) =0$. If we do that, then the last two terms are automatically zero and we have:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 - z(t)\ m\ d^2x(t)/dt^2\ ]$$

Now we said that the path x(t) was a minimum of the action. At a minimum of any function, F(x), the change in the function for a small change in x is zero. Therefore the action must not change when we make the small change of shifting by z(t), *for any z(t)*!!! Hence, it must be true that:

$$0 = m\ d^2x(t)/dt^2$$

This is Newton's equation, and we see that it is just the statement that the action is minimal (or more generally, extremal) for the true physical trajectory, x(t).

Now comes Emmy Noether. Suppose that we make a uniform shift in our coordinate system. This means that we replace out trajectory x(t) by the new trajectory x(t) + z, where z is a constant. Now, even the endpoints of the trajectory, x_0 and x_1, must shift to $x_0 + z$ and $x_1 + z$.

"The physics cannot depend upon such a shift"; this is the principle of translational invariance. What does this statement mean? It means that the action for the trajectory x(t) must be the same as for the trajectory x(t)+z. Well, we have already done the work in computing the shift in the action for any z(t) above. We know that under such a shift the action changes into:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 - z\ m\ d^2x(t)/dt^2\ ] + mv(t_1)\ z - mv(t_0)z$$

and we already know that Newton's equation must hold, so we find:

$$\int dt\ [(1/2)m(dx(t)/dt)^2\ ] + mv(t_1)z - mv(t_0)z$$

The statement that the action cannot change with our shift therefore requires that:

$$0 = mv(t_1) - mv(t_0)$$

This says that there is a quantity, called momentum, given by mv(t) = m dx(t)/dt which must be the same at time t_1 as it was at time t_0.

The statement that the Action cannot depend upon a translation of the coordinate system implies that there is a conserved quantity, called momentum.

Voila!!! Noether's theorem!

This gives you an idea of how it works in the simplest case. Now for some hand-waving. It can easily be generalized to any number of free particles, and you would find that the total momentum is conserved. To include forces requires a significant new idea, however. Instead of integrating the energy to define the action, we integrate something called the "Lagrangian". The energy of a single particle in a potential V(x) is $(1/2)m(dx(t)/dt)^2 + V(x)$. The Lagrangian is the the difference between the kinetic and potential energy:

$$L = (1/2)m(dx(t)/dt)^2 - V(x)$$

The action is defined as the time integral of the Lagrangian:

$$\int dt\ [(1/2)m(dx(t)/dt)^2 - V(x(t))]$$

and the action principle now states:

The Action Principle: The action is the time integral of the Lagrangian. The system moves along the particular trajectory that has an extremal value of the action.

Now again we can choose a particular trajectory, x(t), which extremalizes the action, and again we can add a small shift z(t), and repeat our derivation as before, and we will again obtain Newton's equation, but now in the form:

$F = m\ d^2x(t)/dt^2$, where F is the force, F = -dV/dx.

The momentum will change for the single particle in the potential, because we are not allowing translational invariance anymore, i.e., the potential is fixed in space, and as far as the single particle is concerned, there is a prefered place in space, e.g., the minimum of the potential, etc. We thus find:

$mv(t_1) - mv(t_0) = \text{Impulse}$

where the impulse is just:

$$\int dt\, F(t)$$

If we consider N particles, and have an arbitrary potential interaction between them,

$V(x_1, x_2\ldots , x_N)$

and we furthermore demand that the potential is translationally invariant:

$V(x_1, x_2, \ldots , x_N) = V(x_1+z, x_2+z, \ldots , x_N+z)$

Then we will find, indeed, that the total momentum of all the particles is conserved. All of this follows by the same kinds of manipulations we did above for the free particle.

We can prove energy conservation by shifting the time endpoints of the action integral by a small amount, and demanding that nothing change. We can prove angular momentum conservation by performing a small rotation of all the rotatable things in the action. We use action principles to define field theories, and we get all of the usual conservation laws from them, and more (such as electric charge, and quark color, etc.).

By the way, Feynman, following Dirac, formulated Quantum Mechanics in, perhaps, its most elegant and useful way. He said that a quantum system evolves along all possible trajectories, each trajectory having an amplitude given by

Amplitude for trajectory x(t) = exp[i Action(x(t)) / h-bar]

To compute how the system actually does evolve, we are told to sum (or integrate) over all possible paths (``path" means "trajectory"). Then, the probability of the system getting to the final point on the trajectory is just the (absolute) square of the total amplitude. This is called the Path Integral.

Isn't it spooky? A quantum particle doesn't follow just one trajectory; rather it follows all possible trajectories, each having its own "amplitude". If you block off some of the virtual trajectories the particle might take, you affect the probability that it arrives at its destination! Because Feynman's Path Integral is Quantum Mechanics, and because it involves the Action, the symmetries lead to conservation laws ala Noether, just as they do in Classical Physics!

That's all for now folks (writing equations in HTML is really a pain; but over time we'll add more to this --- CTH)