The Cosmic Asymmetry Between Matter and Antimatter

10

It seems the universe today is almost entirely matter. Evidence from both cosmology and particle physics (the study of the universe on the largest scale and the smallest) now suggests an explanation.

• • •

Frank Wilczek

All the fundamental constituents of matter come in matched pairs: for every kind of particle there is an antiparticle that is identical in mass but opposite in other properties, such as electric charge. The symmetrical pairing of particles and antiparticles is required in order to unite the two great theories of 20th-century physics: relativity and quantum mechanics. The symmetry has been well verified by experiment. Since 1932, when the positron, or antielectron, was discovered, the catalogue of antiparticles has grown apace with the catalogue of ordinary particles. Indeed, a particle and its antiparticle have often been discovered simultaneously when the two were created as a pair by a high-energy collision in a particle accelerator. Such collisions always seem to yield matter and antimatter in equal quantities; indeed, it was long assumed that the laws of nature express no preference for matter or antimatter.

And yet in the world outside the laboratory antimatter is almost never encountered. The atoms composing the earth consist of neutrons, protons and electrons, but never their antiparticles. Does this asymmetry prevail throughout the universe? That is, does the entire universe consist predominantly of matter, with very little antimatter? If it does, has the asymmetry always existed, or did the universe begin with equal numbers of particles and antiparticles and somehow develop an imbalance later?

Recent findings in cosmology and particle physics suggest answers to these questions. They suggest that in the first instant after the big bang, when the universe was much hotter and denser than it is now, there were equal amounts of matter and antimatter. Before the universe was 10^{-35} second old, however, violent collisions among particles created conditions that led promptly to an asymmetry between matter and antimatter. The asymmetry has been locked into the universe ever since. The road leading to this conclusion is still unpaved in places, but I shall try to show that the route is the right one.

How can one be sure the universe consists entirely of matter? It is easy to demonstrate that matter and antimatter cannot be mixed homogeneously. Whenever a particle and the corresponding antiparticle come together, they annihilate each other and their mass is converted into energy. Hence a star made up of half matter and half antimatter would immediately disappear in a titanic explosion. The possibility remains, however, that matter and antimatter might coexist in the universe if each was confined to isolated regions separated by empty space.

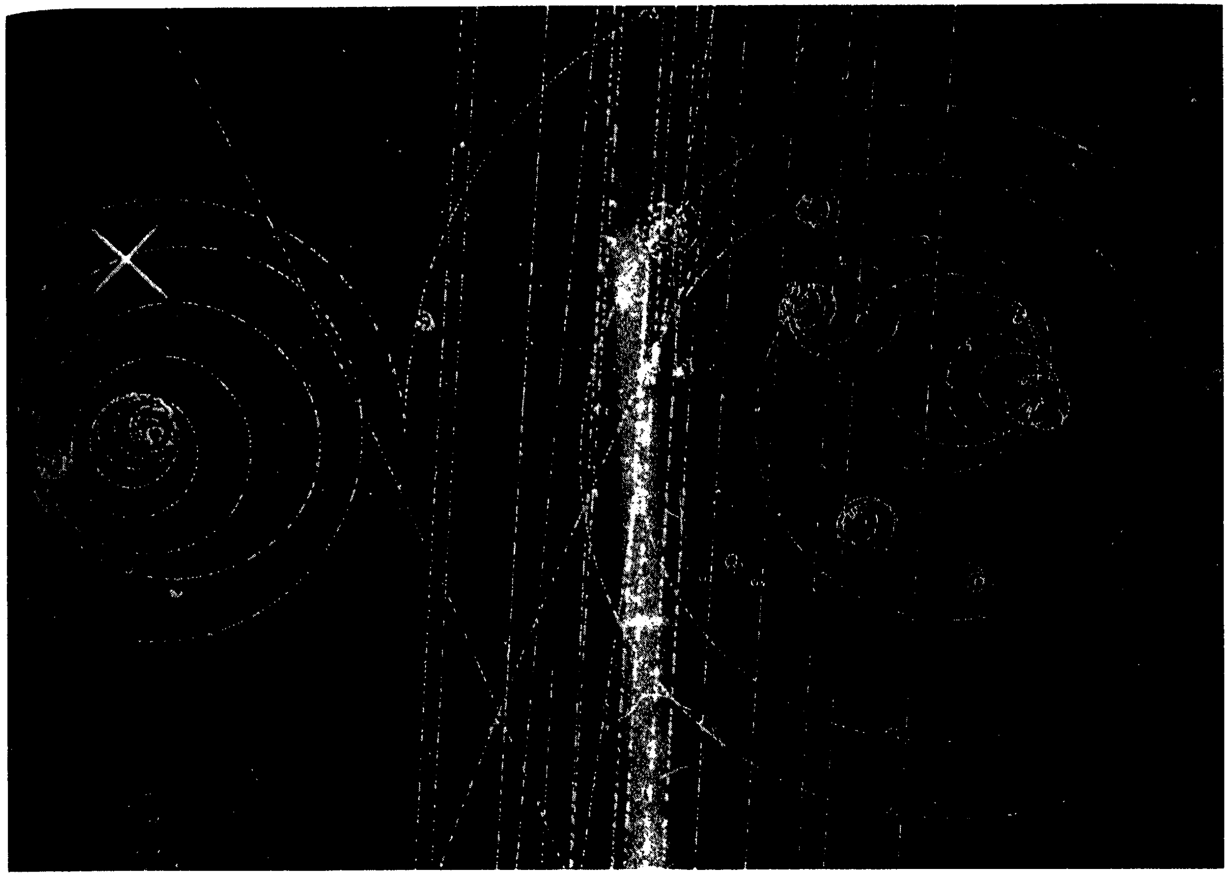


Figure 73 CREATION OF ANTIMATTER, here a positron, becomes visible in a bubble chamber where the trajectory of any electrically charged particle with an electric charge is marked by a trail of bubbles in liquid helium. The positron, whose clockwise spiraling path in the chamber magnetic field fills the right two-thirds of the photo, is the antiparticle of the electron, whose path is the smaller counterclockwise spiral. The electron-positron pair was created by a photon that is invisible because it has no charge. (Photo by Nicholas P. Samios.)

One line of evidence for the preponderance of matter over antimatter is provided by cosmic rays, the high-energy particles that arrive from space. They seem invariably to be particles of matter such as protons and electrons and atomic nuclei made up of protons and neutrons; antiparticles are not observed. Although the origin of cosmic rays is not yet fully understood, they certainly come from throughout the galaxy, and some of them may have a still more distant origin. It therefore seems established that the Milky Way consists entirely of matter, and it is only a little less certain that the group of galaxies of which the Milky Way is a member is also all matter.

Ascertaining that more distant galaxies are composed of matter is a more difficult problem. Merely looking at a galaxy offers no hint of whether it is made up of matter or antimatter. "Looking at" a galaxy implies the detection of photons, or quanta of electromagnetic radiation. The photons include not only those of visible light but also those of radio waves, X rays, gamma rays and so on. The problem is that the photon is its own antiparticle, and there is no way to distinguish a photon emitted by matter from one emitted by antimatter. As a result the light from an antimatter galaxy would be identical with that from a matter galaxy, even in the detailed structure of the spectrum. For example, the characteristic emission lines of the hydrogen atom would be duplicated exactly in emission lines of the antihydrogen atom.

There is one circumstance in which photon observations might indirectly reveal the presence of antimatter. If an antimatter galaxy were close to a matter galaxy, the boundary region between them would be the site of frequent particle-antiparticle annihilations. The energy of each such annihilation would eventually appear in the form of photons at gamma-ray wavelengths. The border region would therefore be a place where gamma radiation is copiously emitted. Astronomical sources of gamma radiation are known and are under investigation, but no source with the proper characteristics has been found. This argument is of no consequence, however, if empty space separates the matter from the antimatter. At best the failure to observe strong gamma emissions suggests that clusters of galaxies must consist entirely of matter or entirely of antimatter, not a mixture of the two. The clusters are pervaded by intergalactic gas, and any difference in composition within a cluster would give rise to gamma radiation.

In the future the question of whether any substantial aggregations of antimatter exist in the universe may be answered by the advent of telescopes that detect not photons but neutrinos. Unlike the photon, the neutrino has a distinguishable antiparticle. Neutrinos and antineutrinos would be emitted in different proportions by nuclear reactions in matter and antimatter. A star composed of matter radiates mainly neutrinos, whereas a star composed of antimatter would give rise chiefly to antineutrinos. The issue has not been settled yet by neutrino observations because building a neutrino telescope is a formidable project. Neutrinos have negligible mass and hardly interact at all with other matter; their detection is problematic.

For now at least, the prevailing opinion among astronomers and astrophysicists is that matter dominates over antimatter in the present universe. As I have suggested, the evidence in support of this view is not compelling, although there is a notable lack of evidence for the existence of antimatter. What ultimately seems decisive is the difficulty of imagining how matter and antimatter in the early universe could have become segregated into distinct regions. It seems more likely they would have simply annihilated each other everywhere.

If the universe is now mostly matter, one is moved to ask how this asymmetry came about. One possibility is that the preference for matter was built in at the start, that the primordial material issuing from the big bang was predominantly matter. This hypothesis cannot be disproved, at least for now, but it is rather unsatisfying. Virtually any composition of the universe could be explained in the same way. Moreover, the primordial-imbalance hypothesis accords fundamental status to a set of initial conditions that have no apparent rationale; any number of alternatives seem equally plausible. If a theory consistent with established physical principles could be constructed in which the universe was initially symmetrical, it would be more appealing. It is just such a theory that is offered by the conjunction of cosmology and particle physics.

A crucial event in modern cosmology was the discovery in the 1920's by Edwin P. Hubble that distant galaxies are receding from the earth with speeds proportional to their distances. The recession of the galaxies implies that the entire universe is expanding. Extrapolating backward in time leads to the conclusion that roughly 10 billion years ago the material that now forms the galaxies emerged explosively from a highly compressed state. Indeed, following the backward evolution to its mathematical limit suggests that the entire universe was initially a dimensionless point.

At the instant of the big bang the density and the temperature of the universe were infinite. The temperature fell rapidly, but throughout the first minute it was greater than 10^{10} degrees Kelvin. Under those conditions any atoms that may have formed were immediately torn apart; even atomic nuclei could not survive but were decomposed into their constituent particles. In other words, the universe in its first moments was a hot plasma of free particles, many of which, such as the electrons and the protons, were electrically charged. Because charged particles in motion give off electromagnetic radiation, the early universe was rich in photons.

The expanding universe cooled much as an expanding gas cools, and by about three minutes after the big bang protons and neutrons began to combine to form the nuclei of helium atoms. The remaining unbound protons would eventually become hydrogen nuclei. (All the heavier elements, which are quite rare on a cosmic scale, have been built up out of hydrogen and helium in the cores of stars and in supernova explosions.) By making the simplest assumptions about the conditions in the early universe that are consistent with known physical laws, one can calculate that the ratio of helium to hydrogen was about one to three by weight. The value is in good agreement with the ratio estimated for the universe today. The success of this prediction is testimony to an understanding of what the universe was like a few minutes after its birth.

After roughly 10,000 years of expansion the universe was cool enough for the last of the free charged particles to be incorporated into atoms. Each atom is electrically neutral because it has equal numbers of positive and negative charges. Photons interact only weakly with neutral matter, and so from that time forward the matter and the electromagnetic radiation in the universe were essentially uncoupled. Since then the radiation has freely followed the expansion of the universe, cooling all the while. How can radiation cool, and how can it have a temperature in the first place? If the radiation is regarded as a gas of photons, then it cools by expansion, somewhat like a gas of material particles, as the average energy of the photons decreases. If the radiation is regarded as a wave, then the expansion of space brings an increase in the distance between any two successive wave crests. The longer wavelength corresponds to a smaller photon energy.

In 1964 it was discovered that microwave radiation is striking the earth evenly from all directions. The radiation corresponds to a photon gas that fills the universe to a density of about 300 photons per cubic centimeter. The temperature of the radiation is 2.7 degrees K., a value much reduced from the temperature of about 10,000 degrees at the time of decoupling. The presence of the radiation is further evidence that this theoretical reconstruction of the early universe is correct. Emboldened by these successes, one can attempt to extrapolate back to the earliest moments of the universe to see if the extreme conditions then prevailing might account for the present asymmetry between matter and antimatter.

In the first few seconds of the universe the particles of the hot primeval gas had an average energy that exceeds the capabilities of even the largest modern particle accelerators. Interactions of particles at those energies may have been qualitatively different from all those that can be observed now. Even if the events in the early universe differed in character from those accessible today, however, the laws of nature governing the events can be assumed to endure unchanged. What is needed, then, is a theory that will predict how particles act at very high energy on the basis of natural laws deduced from events at much lower energies.

Among the natural laws in question are conservation laws applied to quantum numbers. A quantum number is essentially a bookkeeping convenience, adopted as an aid to keeping track of the various properties of particles. For example, electric charge can be expressed as a quantum number: the proton is assigned a value of $+1$, the electron a value of -1 and the photon and all other neutral particles a value of zero. The conservation law that applies to electric charge states that the total charge quantum number cannot change in an interaction; the sum of all the charge quantum numbers after the event must be equal to the sum before the event.

It is important to note that the conservation of electric charge does not forbid a change in the number of charged particles. An electron and a positron can annihilate each other, diminishing the number of particles by two; the total charge, however, is zero both before the annihilation and after it. The opposite process, in which an electron and a positron are created out of pure energy, obeys the conservation law for the same reason. Indeed, any particle can be created or annihilated simultaneously with its antiparticle, and all quantum numbers will automatically be conserved.

A quantum number called baryon number is of notable interest in tracing the source of the cosmic asymmetry between matter and antimatter. The baryons are a large family of particles whose most familiar members are the proton and the neutron; as basic constituents of atomic nuclei, the baryons clearly have an important role in the structure of ordinary matter. The proton, the neutron and all the many related baryons are assigned a baryon number of $+1$. For the antiproton, the antineutron and other antibaryons the baryon number is -1. All other particles, including the pions, the muons, the neutrinos, the electron, the photon and their antiparticles, have a baryon number of zero.

The conservation of baryon number is the assertion that in any reaction the baryon number of all the particles in the initial state is equal to the baryon number of all the particles in the final state. Again the number of particles can change, as when a proton and an antiproton are created or annihilated as a pair, but the net baryon number remains unaltered. Suppose, for example, two protons (with a total baryon number of +2) collide at high energy. The final products might include four protons, a neutron, three antiprotons and a number of pions; adding up the baryon numbers shows that the total remains +2.

Electric charge is a quantity that is thought to be conserved under all circumstances. The absolute conservation of baryon number is less certain, and indeed there is now strong suspicion that the law is occasionally violated.

The most compelling evidence for the conservation of baryon number is the stability of the proton. As the least massive particle whose baryon number is +1, the proton cannot decay into any set of lighter particles without violating the conservation law. Detection of a proton decay would therefore constitute direct evidence that the law is not always enforced.

No one has yet seen a proton decay, and even crude calculations suggest that its lifetime is long. If protons decayed, for example, in human bone, the energy released would increase the incidence of cancer. On this basis the lifetime of the proton must be greater than 10^{16} years: If protons decayed on Jupiter, the energy would contribute to the luminosity of the planet. On this basis the lifetime is greater than 10^{18} years. Systematic experiments suggest that the lifetime is actually greater than 10^{29} years. In contrast, the age of the universe is only 10^{10} years. Evidently if the proton does decay, it is an exceedingly rare event. If the actual lifetime should turn out to be 10^{30} years, then in 100 tons of matter (a sample of 10^{31} protons) an average of 10 would decay in a year. The low rate suggests both the stringency of the law of conservation of baryon number and the difficulty of mounting experiments to search for violations. Several such experiments are nonetheless under way.

Saying that the universe has an excess of matter over antimatter is equivalent to saying that it has a positive baryon number. If the law of conservation of baryon number were absolute, the number would have been constant through the eons. There may have been more of both baryons and antibaryons once, but the number of baryons minus the number of antibaryons would have always been the same.

Consider the state of the universe when it was a hundredth of a second old and had a temperature of 10^{14} degrees K. For any given temperature there is an equilibrium mixture of different kinds of particles such that for each kind the number of particles being created by collisions or decays balances the number being destroyed. In the early universe, at 10^{14} degrees, the equilibrium mixture included about a billion protons and a billion antiprotons for every proton in the present universe. If the baryon number of the universe was the same then as it is now, the ratio of protons to antiprotons must have been roughly 1,000,000,001 to 1,000,000,000, and so the asymmetry would have been scarcely noticeable.

Later almost all the protons were annihilated by encounters with antiprotons. Only the conservation of baryon number forestalled a total annihilation of all baryons and antibaryons. In this view all the present protons, and therefore all the present galaxies, stars, planets and sentient beings, are the residue of a one-part-in-a-billion imbalance. It is the small imbalance, the early manifestation of the cosmic asymmetry between matter and antimatter, that stands in need of explanation. Once the excess of matter has been established the subsequent evolution of the universe is comparatively straightforward; the source of the original asymmetry is a deeper mystery. In particular, if the universe evolved from an initial state that was fully symmetrical between matter and antimatter (a state having a baryon number of zero) into an asymmetrical state in which the baryon number is greater than zero and protons outnumber antiprotons, then the conservation of baryon number must have been violated at some stage.

The first indication that the conservation of baryon number cannot be exact came from a distantly related field of inquiry: the theory of the black hole. A mathematical analysis demonstrated that the only properties of a black hole measurable by an outside observer are its mass, its angular momentum and its electric charge. Notably absent from the list is the baryon number. Hence a black hole created by the collapse of a star would be indistinguishable from one created by the collapse of an antistar with the same mass, angular momentum and charge. Yet the baryon number for the star is positive, whereas the number for the antistar is negative. Clearly there is no way to assign a baryon number to a black hole and be certain that the baryon number of the universe is conserved.

The putative violation of the conservation law by black holes suggests that a similar mechanism on a microscopic scale might lead to proton decay. In this hypothetical process a proton is absorbed by a virtual black hole: a minute, short-lived fluctuation in the geometry of space-time, which in principle could arise anywhere and at any time. The virtual black hole promptly decays into a positron and a gamma ray. In these particles the mass or energy of the proton reappears and so does its positive electric charge; its baryon number, however, is irretrievably lost. Although the details of the hypothetical process are uncertain, estimates suggest that it implies a lifetime for the proton on the order of 10^{40} years. If the conservation of baryon number is violated in this way, the violation is feeble indeed.

A second indication that the conservation of baryon number is only approximate is slightly less exotic and also more powerful in its effect on the lifetime of the proton. This second possible mechanism is an outcome of revolutionary developments in the theories describing interactions among elementary particles. To be specific, it is an outcome of the understanding, achieved only in the past decade, that the '"strong" force responsible for holding together the nuclei of atoms and the "weak" force responsible for most radioactive decays are quite similar to electromagnetism.

How could improved understanding of these forces, which do not violate the conservation of baryon number, lead to theories predicting such a violation? A more detailed discussion of the forces must precede the explanation.

Of the three forces only electromagnetism is routinely evident in the macroscopic world that people perceive directly. The electromagnetic force acts only between particles that have an electric charge; the interaction can be described as the exchange of a third particle, namely a photon. The photon is said to be a vector particle, a designation given to any particle whose spin angular momentum, when measured in fundamental units, is equal to 1. Perhaps the most fundamental characteristic of electromagnetism is that it can be described by a gauge-invariant theory. In a theory of this kind the origin of the force is related to a conservation law, in this case the conservation of electric charge. The coupling of vector particles to a conserved charge is characteristic of gauge theories (see Chapter 5, Figure 38).

In all these respects the strong interaction is similar. The force arises from a gauge theory, and a strong interaction can be described as the exchange of a vector particle by two other particles that have a certain kind of charge. The vector particle is not the photon, however, but a hypothetical entity called a gluon, and the charge is not electric charge but a property called color. The color charge of course has nothing to do with color in the ordinary sense. The word charge in this context is less fanciful. The word is apt because color charge plays much the same role in the strong interaction as electric charge does in the electromagnetic interaction.

One difference between electromagnetism and the strong interaction is that electromagnetism has only one kind of charge, whereas in the strong interaction there are three, labeled *R*, *G* and *B* for red, green and blue. The colors are carried by the funda-

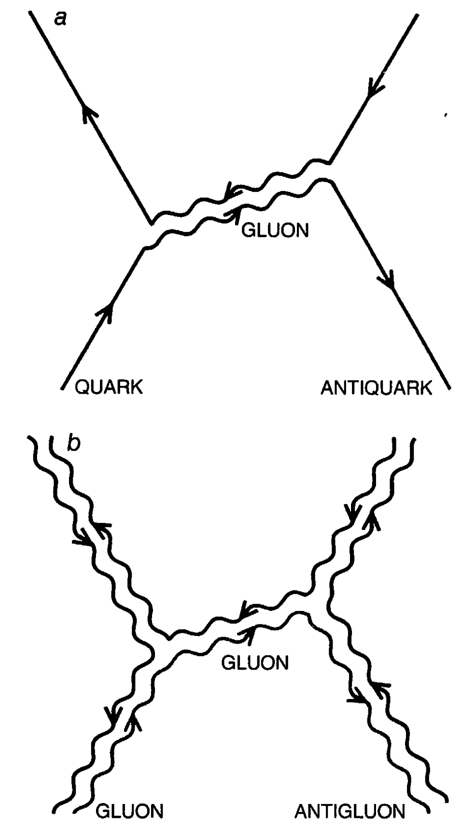


Figure 74 STRONG FORCE can be represented as the exchange of a gluon between two particles that have color. In *a* the particles are quarks; the one on the left is blue, the one on the right is antiblue. The strong interaction changes the trajectory and the color of each quark. To conserve color throughout the interaction the gluon must have both a color and an anticolor; as a result the gluons themselves are subject to the strong interaction. Gluon and antigluon scattering is shown in *b*.

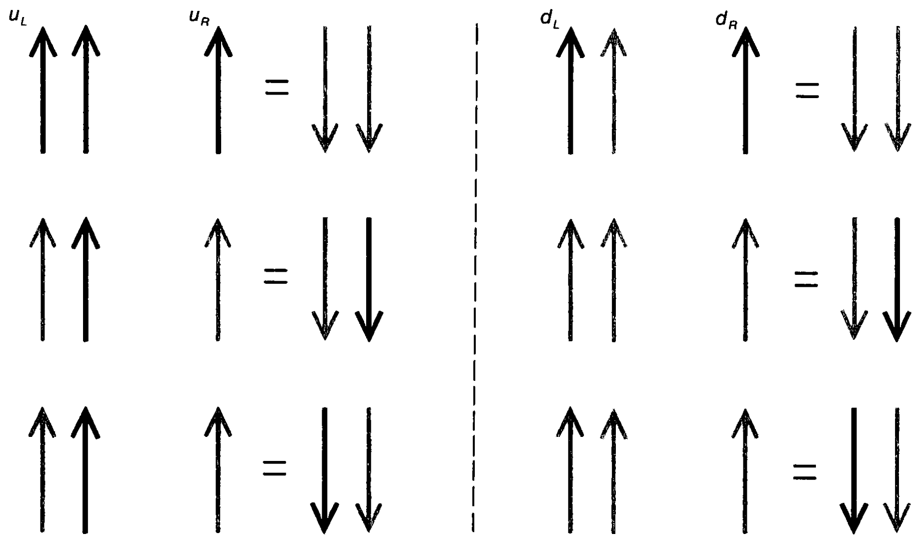


Figure 75 FAMILY OF 12 QUARKS. Each quark has four salient properties: type (e.g., *u* or *d*), color charge with respect to the strong interaction (red, green or blue), spin and for left-handed particles weak charge (purple for *u*, orange for *d*). In the mathematical formalism of the strong interaction a quark with a given color is equivalent to a quark without that color but with two anticolors (shown for right-handed quarks).

mental constituents of all strongly interacting particles: the quarks. Each quark has a single color, denoted by an assignment of the three color quantum numbers. For red quarks R equals $+1$ whereas G equals 0 and B equals 0. Similarly, for green quarks G equals $+1$ and for blue quarks B equals $+1$ and the other color quantum numbers are zero. Eight kinds of gluon are required by the theory. Six kinds change a quark of one color into a quark of a different color in all possible ways, namely red into green, red into blue, green into red, green into blue, blue into red and blue into green. The other two gluons resemble the photon in that they carry a force between "charged" particles but do not alter the charge.

In addition to color quarks have three other properties. Each quark is called up (*u*) or down (*d*). The *u* quarks have an electric charge of $+2/3$, the *d* quarks an electric charge of $-1/3$. Each quark also has a spin whose axis is aligned with the particle's direction of motion (subscript R for right-handed particles), or is opposite to that direction (subscript L for left-handed particles). Finally, the left-handed quarks have a color with respect to the force in nature called the weak interaction, which mediates most radioactive decays (purple for *u*, orange for *d*).

A property of color charges is that they can cancel one another. For example, the combination of one red, one green and one blue quark is a colorless composite particle, to which gluons do not couple. (Similarly, particles with opposite electric charges can combine to form a neutral composite.) It is only such colorless combinations of quarks that seem to appear in nature. All baryons consist of three quarks, one quark in each of the three colors. The mesons, which make up another category of strongly interacting particles, each consist of a quark and an antiquark.

A second difference between the strong interaction and electromagnetism is that the gluons themselves are charged, whereas the photon is not. For example, the gluon that is absorbed by a red quark and transforms it into a green quark has R equal to -1, G equal to $+1$ and B equal to 0; with this combination of colors and anticolors color charge is conserved throughout the interaction. Since gluons couple to colored particles and since gluons themselves are colored, gluons couple to one another. In contrast, the photon is electrically neutral and does not couple to other photons. The difference has a profound dynamical consequence: at short distances the strong interaction loses strength. Quarks bind only feebly when they are close together, but their binding becomes quite powerful when they are somewhat farther apart. (In the present context a long distance is 10^{-13} centimeter.)

This paradoxical force law explains a great deal. It has been known since the mid-1960's that the properties of strongly interacting particles could be accounted for by the quark model, but no one has ever observed an isolated quark. Furthermore, the utility of treating a strongly interacting particle as a composite of quarks rests on an approximation in which the quarks are essentially noninteracting particles inside a communal "bag." It was puzzling that strongly interacting particles such as quarks could successfully be described as noninteracting. The notion that the strength of the strong interaction among quarks decreases when the quarks are close together neatly explains why the quarks inside a "bag" interact only feebly with one another and yet cannot be pulled far apart. It may be impossible to isolate a quark. The gauge theory of the strong interaction that underlies the quark model leads to many experimental predictions, which so far have proved very successful. The theory is gaining almost universal acceptance.

The weak interaction can be described in much the same way as the electromagnetic and the strong interactions, but it has a few twists of its own. First, there are two kinds of charge, analogous to the three color charges of the strong interaction. I shall call them P and O, for the colors purple and orange. Three vector particles, called W^+, W^- and Z, mediate the interaction. These particles have large masses, unlike the photon and the gluons, which are massless. A particle with a large mass can arise spontaneously only as a short-lived fluctuation; if it is short-lived, it cannot go far, and as a result the weak interaction has a very short range. A more surprising characteristic of the weak force is that it acts only on particles with certain geometric properties. Quarks, electrons, neutrinos and a few other particles can be classified as right-handed or left-handed according to the relative orientation of their spin angular momentum and their linear motion. A right-handed particle has its spin axis pointing parallel to its direction of motion, a left-handed particle antiparallel. The weak interaction affects only left-handed particles and right-handed antiparticles. In sum, the strong and the weak interactions require five kinds of color charge (red, green and blue for the strong and purple and orange for the weak), along with vector particles that transmute some of these colors.

In the theories I have outlined here the strong force is a mechanism for changing the red, green and blue colors of quarks. The weak force works similar changes on the purple and orange color quantum numbers of particles. If these theories are to be truly unified, one would expect some additional force to transform the strong colors into weak colors and vice versa. In addition to being aesthetically attractive, a scheme that incorporates such a new force accommodates all known particles quite neatly. Moreover, it makes definite predictions. For example, it predicts the mass of the W.

It is by postulating a new force that the unifying theories compromise the conservation of baryon number and allow the proton to decay. New color-changing vector particles are introduced as bridges between particles with strong color, such as the quarks composing a proton, and particles with only weak color, whose baryon number is zero. I shall designate such vector particles X. The unifying theory predicts that the X has a mass that is 10^{15} times the mass of the proton (and is roughly comparable to the mass of a flea), compressed into a volume only 10^{-27} centimeter across. Because the X particle is so massive, its spontaneous creation is extremely rare. Accordingly it is estimated that the mean lifetime of the proton is long but not infinite; the lifetime should be on the order of 10^{31} years.

To be sure, a lifetime of 10^{31} years implies that in the universe today the violation of the conservation of baryon number is slight. As I have noted, however, the matter-antimatter asymmetry observed

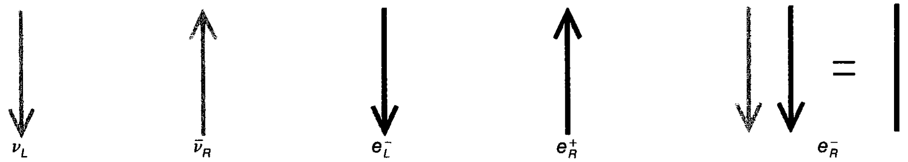


Figure 76 WEAKLY INTERACTING PARTICLES include the left-handed neutrino; its antiparticle, the right-handed antineutrino; the left-handed electron, and its antiparticle, the right-handed positron. These particles do not interact strongly and so they are shown without strong color charges. The right-handed electron (*right*) is given two weak color charges, a configuration equivalent to a state with no weak color charge (*black line without arrow*).

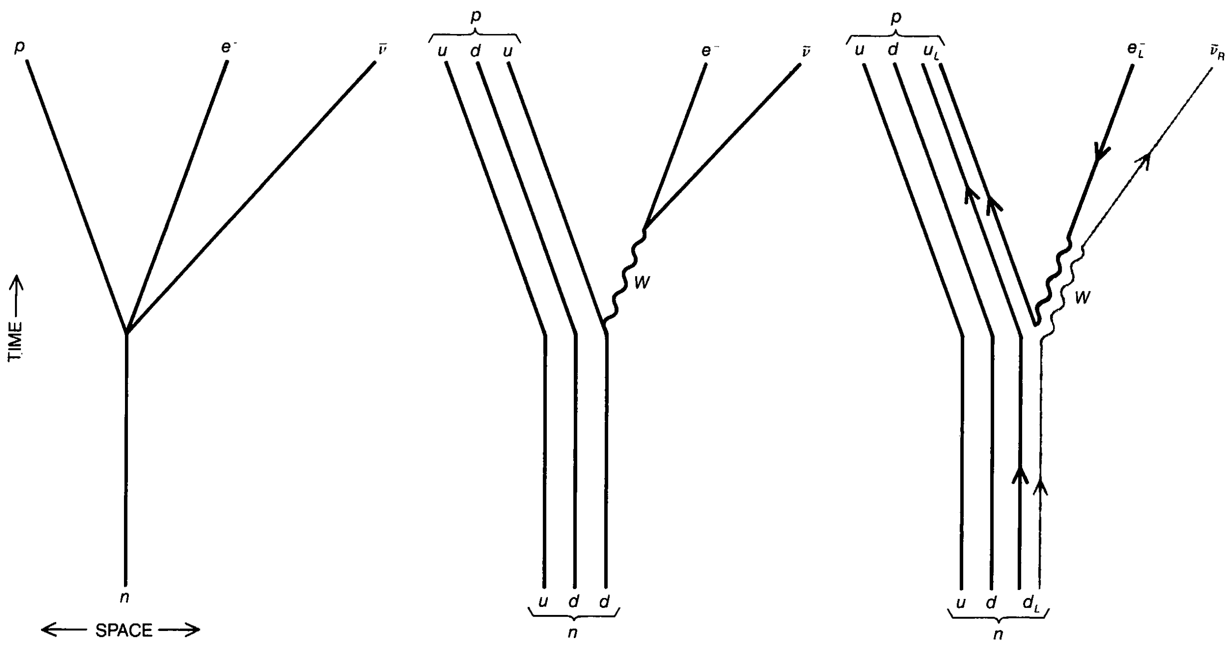


Figure 77 DECAY OF THE NEUTRON by a weak interaction averages 15 minutes. In the broadest view (*left*) the decay transforms the neutron (n) into a proton (p), an electron (e^-) and an antineutrino ($\bar{\nu}$). In a finer analysis (*center*) the neutron consists of three quarks and only a d quark, arbitrarily placed at the right, is affected by the decay. A left-handed d quark (*right*) decays into the left-handed u quark, the electron and the antineutrino. The orange charge of the d_L quark is transformed by a short-lived (or virtual) particle W, which is orange and antipurple. Only weak colors are changed.

today corresponds to merely a one-part-in-a-billion asymmetry in the early universe. Moreover, a mode of decay that requires the creation of an unstable heavy particle may well have been commoner in the earliest moments of the universe, when heavy particles could be freely created by ultrahigh-energy collisions.

I turn now to the idea that physical laws are indifferent to the distinction between matter and antimatter. The history of the idea is a series of upset expectations. Until the mid-1950's it was generally thought the laws of physics would remain unchanged if experiments were repeated in a mirror-reflected world. In other words, it was thought no absolute distinction could be made between left and right. A variety of experiments then revealed, however, that mirror-reflection symmetry is badly broken by the weak interactions. An example is provided by the decay of the muon into an electron, a neutrino and an antineutrino. In more than 999 decays in 1,000 the electron is found to be left-handed: its spin axis points in the direction opposite to its direction of motion. Thus the decay of the muon furnishes an absolute standard of left v. right.

Theorists next proposed a more comprehensive symmetry that seemed to be respected by all interactions. This second hypothesis was that the laws of physics would be unchanged by the mirror reflection of an experiment if at the same time all the particles in the experiment were replaced by their antiparticles. The symmetry is called CP for charge conjugation and parity, or mirror reflection. CP symmetry predicts that in the decay of the antimuon a positron should emerge instead of an electron and the positron should almost always be right-handed. In the case of muon decay exact CP symmetry is observed.

If CP symmetry were absolute, a preponderance of matter or of antimatter could not evolve from a primordial equality between the two. For every process that creates a particle, an equally likely mirror process would create the antiparticle.

The concept of absolute CP symmetry survived for about seven years. Then it was observed that the long-lived neutral K meson, which is its own antiparticle, decays more often into a negative pion, a positron and a neutrino than it does into a positive pion, an electron and an antineutrino. If CP were an absolute symmetry, the two decay modes would have to be equally likely. No violation of CP symmetry has been found except in K-meson decay, but such violations might have a more prominent role in nature at ultrahigh energies.

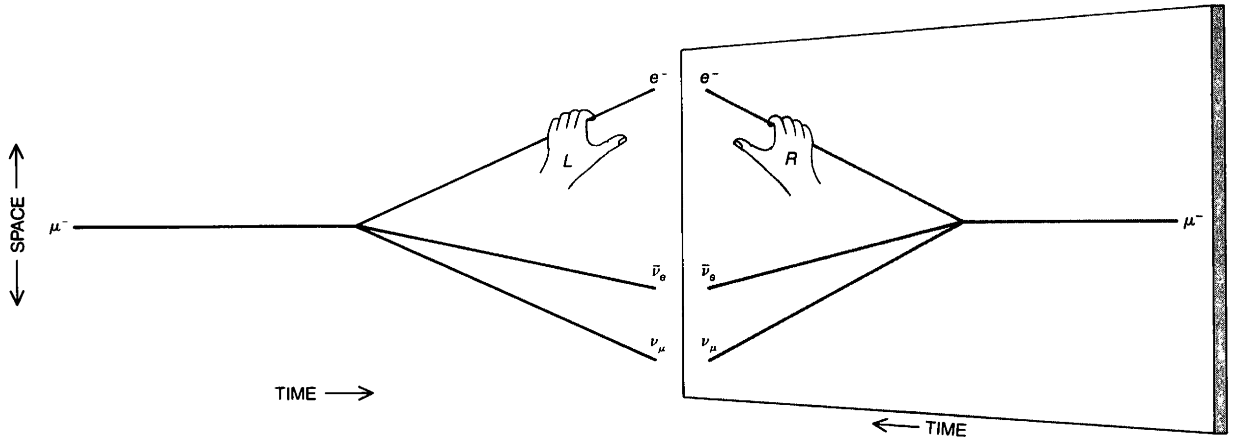


Figure 78 VIOLATION OF PARITY SYMMETRY. Parity, or P, conservation holds that processes remain invariant when they are transformed by a mirror image. The process shown is the decay of a muon (μ) into an electron (e^-), an electron-type antineutrino ($\bar{\nu}_e$) and a muon-type neutrino (ν_μ). The electron is left-handed. In the mirror reflection of the decay the electron is right-handed. In reality parity symmetry is broken in this process: left-handed electrons appear more than 1,000 times as often as right-handed ones.

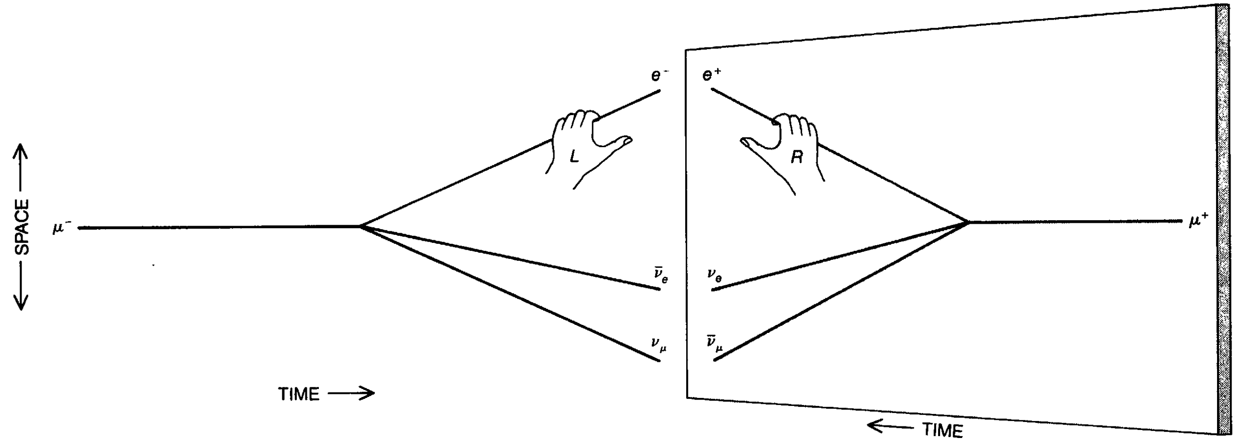


Figure 79 CP SYMMETERY proposed a symmetry that might be observed even if parity symmetry is violated, asserting that the symmetry broken by mirror reflection could be restored by replacing all particles with their antiparticles. In muon decay CP symmetry holds true: decay at the left and right appear to be equally common.

The developments I have described suggest that both the permanence of certain particles, as formalized in the law of conservation of baryon number, and the indifference of physical laws to the distinction between matter and antimatter, as for-

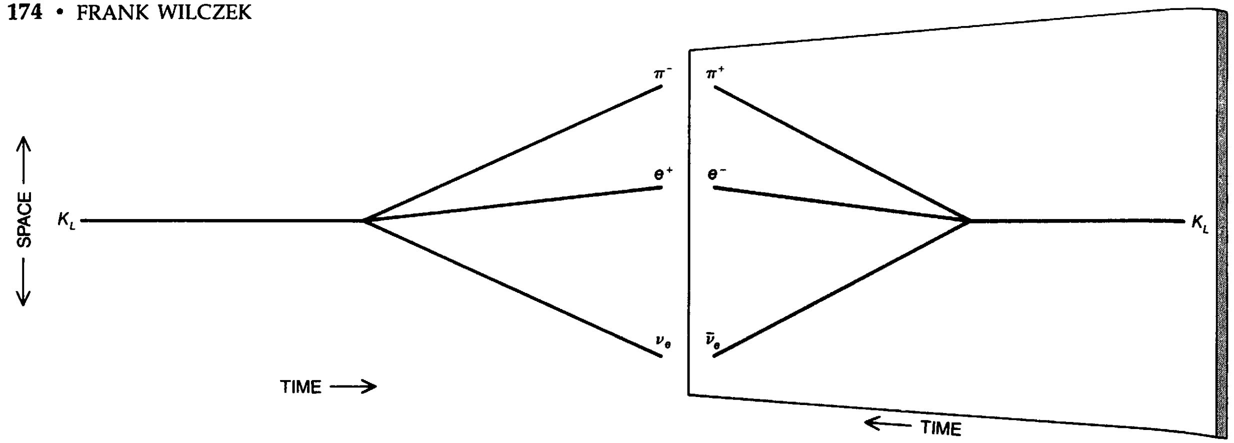


Figure 80 CP VIOLATION was demonstrated in the decay of the long-lived neutral *K* meson (K_L). The decay of this particle into π^-, e^+ and ν_e is commoner than decay into the antiparticles π^+, e^- and $\bar{\nu}_e$. (The long-lived neutral *K* meson is its own antiparticle.) If CP symmetry were never broken, the ratio of baryons to antibaryons would be fixed and no asymmetry could develop between matter and antimatter.

malized in the principle of CP symmetry, are not exact but only approximate. It is true the principles hold quite accurately today, but this may not have been the case in the very early universe. Indeed, given even a small violation of these principles one can construct a specific chain of events leading from a universe in an initial state of symmetry between matter and antimatter to a universe with a preponderance of matter over antimatter.

The chain of reasoning begins with the observation that the temperature of the universe has been falling steadily since the big bang. The higher the temperature, the higher the average speed and energy of the particles that make up the universe, and hence the greater the energy available in a collision for the creation of other particles. At a temperature greater than 10^{28} degrees K. the typical energy of a particle was comparable to the rest-mass energy of an X particle. Until about 10^{-35} second after the big bang the universe had such a temperature, and so one can propose that it had a great density of X particles.

As the universe expanded and cooled, the probability of creating an X particle declined rapidly; meanwhile the existing particles were rapidly decaying. Suppose the decays did not conserve baryon number. An X particle might then decay into any of several final states with differing total baryon number. The average might be, say, $+\tfrac{2}{3}$. If the universe had equal amounts of matter and antimatter before it was 10^{-35} second old, it would include equal numbers of X's and $\overline{X}$'s, where the $\overline{X}$ is the antiparticle of an X. It might seem, therefore, that every decay mode of an X would be counterbalanced by the decay of an $\overline{X}$, which would yield particles with an average baryon number of $-\tfrac{2}{3}$. In that case the total baryon number of the universe would remain zero at all times. Actually, since CP symmetry may not have been observed exactly in the decay of the X and the $\overline{X}$, one cannot conclude that the two decay sequences always yielded symmetrically opposite sets of particles. The $\overline{X}$ might give rise to particles whose average baryon number was not $-\tfrac{2}{3}$ but rather, say, $-\tfrac{1}{3}$.

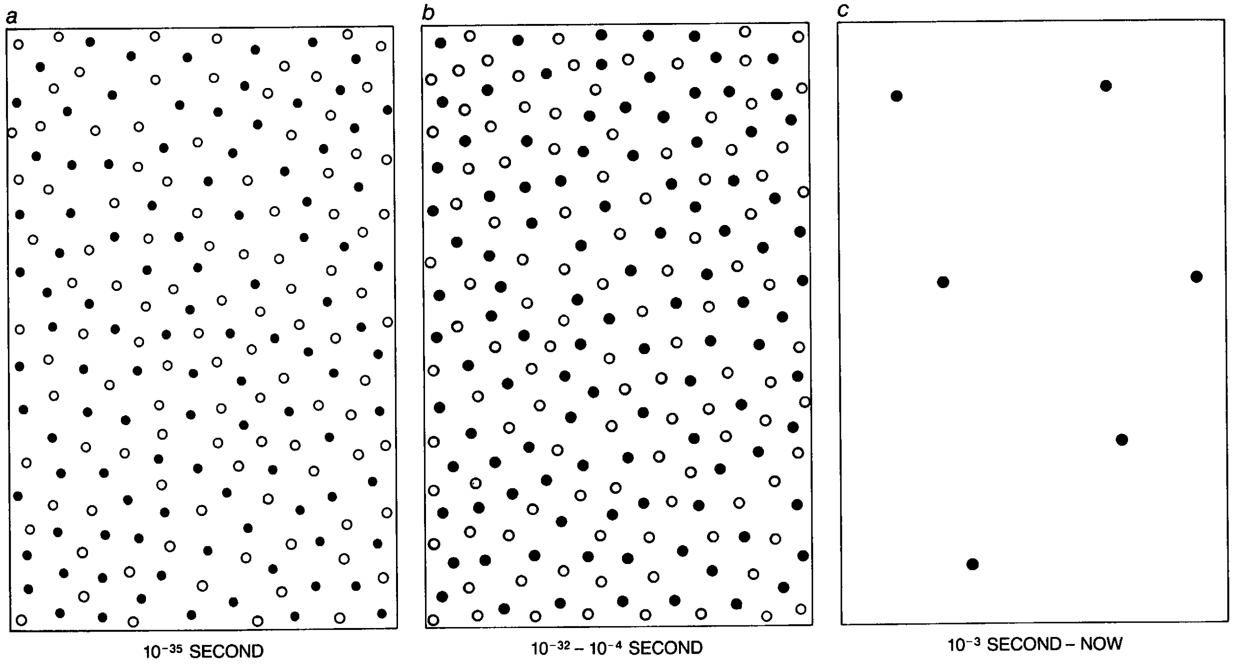
In this way a universe that had equal numbers of X and $\overline{X}$ particles would have evolved into a universe with a positive baryon number and a corresponding preponderance of matter. It could have been a universe, for example, with a one-part-in-a-billion imbalance favoring matter. After the first 10^{-35} second or so the temperature and the typical energy per particle throughout the universe would fall below the threshold for the creation of an X and an $\overline{X}$. The processes that violate baryon number would then become insignificant, and the preponderance of matter over antimatter would be frozen in. The universe would still have many more baryons and antibaryons than it has now, but most of them would eventually annihilate one another, leaving the residue of matter observed today.

Several aspects of this argument are highly speculative, and the explanation of the cosmic asymmetry between matter and antimatter may seem more mythical than scientific. To an extent that is unavoidable, since the extreme conditions of the early universe cannot be reproduced in a laboratory. What distinguishes scientific speculation from myth is its logical consistency and the amenability of at least some of its elements to experimental test. I have described how the inner logic of particle physics leads to unified theories in which baryon number is not conserved, and I have noted that future developments both in neutrino astronomy and in the search for the decay of the proton will test the theories. If these difficult experiments give results consistent with theoretical expectations, they will bring much closer the scientific understanding of a mysterious asymmetry. Even now, calculations carried out in accordance with the unified theories suggest that the average density of matter in the universe today is consistent with the primordial course of events the unified theories imply. Because of uncertainties about the mechanisms of CP violation it is difficult to make the calculations precise, but the qualitative picture is satisfying.

A further question remains. I have described how the universe could have begun with symmetry between matter and antimatter and then have grown asymmetrical. Why was the universe symmetrical in the beginning?

At one level this question can be answered statistically. Even if interactions that violate baryon number were frequent in the early universe, the most likely universal condition, which would be attained at equilibrium before 10^{-35} second, is one in which the amount of matter equals the amount of antimatter. Unified theories therefore enforce initial symmetry automatically; it need not be postulated separately. After 10^{-35} second the decay rates of the *X*'s and $\overline{X}$'s would have been slow compared with the expansion and cooling rate of the universe. Under the condition equilibrium could no longer be attained.

At a deeper level I do not find this explanation fully satisfying. It fails to explain why the universe should have begun in an explosive event. It also



10^{-35} SECOND

10^{-32} – 10^{-4} SECOND

10^{-3} SECOND – NOW

Figure 81 EVOLUTION OF COSMIC AYSMMETRY between matter and antimatter. Panel *a*, the universe 10^{-35} second after the big bang, shows equal quantities of *X* particles (*dots*) and their antiparticles, $\overline{X}$'s (*open circles*). In panel *b* the *X*'s and $\overline{X}$'s have decayed leaving a slight imbalance favoring protons (*colored dots*) over antiprotons (*open circles*). In panel *c* each encounter of a proton and an antiproton has caused the annihilation of both particles, and only the excess protons survive (six in the drawing, one in a billion actually).

fails to explain why the universe is symmetrical in several other ways: it is electrically neutral on the average and it seems to have no net angular momentum.

I shall now describe an idea that may lead to an understanding of these questions. It is by no means well established, but it does suggest a program of research. Indeed, it was the original motivation for my own work on the matter-antimatter asymmetry.

Modern theories of the interactions among elementary particles suggest that the universe can exist in different phases that are analogous in a way to the liquid and solid phases of water. In the various phases the properties of matter are different; for example, a certain particle might be massless in one phase but massive in another. The laws of physics are more symmetrical in some phases than they are in others, just as liquid water is more symmetrical than ice, in which the crystal lattice distinguishes certain positions and directions in space.

In these theories the most symmetrical phase of the universe generally turns out to be unstable. One can speculate that the universe began in the most symmetrical state possible and that in such a state no matter existed; the universe was a vacuum. A second state was available, and in it matter existed. The second state had slightly less symmetry, but it was also lower in energy. Eventually a patch of the less symmetrical phase appeared and grew rapidly. The energy released by the transition found form in the creation of particles. This event might be identified with the big bang. The electrical neutrality of the universe of particles would then be guaranteed, because the universe lacking matter had been electrically neutral. The lack of rotation in the universe of matter could be understood as being among the conditions most favorable for the phase change and the subsequent growth, with all that the growth implied, including the cosmic asymmetry between matter and antimatter. The answer to the ancient question "Why is there something rather than nothing?" would then be that "nothing" is unstable.

POSTSCRIPT

Since my article was written, there have been two important experimental developments. New experiments at high-energy accelerators have dramatically supported and strengthened our faith in what is now called the "standard model" of particle interactions. At the CERN proton-antiproton collider, the $W^{\pm}$ and Z^0 bosons were produced and convincingly identified, and the properties of quark and gluon jets predicted by QCD were confirmed in accurate qualitative detail. The standard model ascribes the strong, electromagnetic and weak interactions to exchange of various gauge bosons—respectively, the color gluons of QCD, the photon, and the *W* and *Z* particles—which are naturally thought of as either responding to, or transforming, five different "color" charges. These separate theories fairly cry out for unification into a single theory that allows all possible transformations among the five colors. Theories embodying this unification invariably lead to violation of the law of baryon-number conservation, an essential requirement for a theoretical explanation of the cosmic matter-antimatter asymmetry.

So much for the good news. The bad news is that despite heroic efforts the one manifestation of unification that might have been accessible in terrestrial laboratories—decay of the proton due to baryon number nonconservation—has not been seen. In fact, the proton lifetime almost certainly exceeds 10^{32} years. More sensitive experiments are mainly limited by the sheer size of the necessary detectors. If in several years not a single proton decays within 100 kilotons of material, both the patience of experimenters and their ability to monitor gigantic chunks of matter become severely strained.

The simplest unified models predict proton lifetimes of about 10^{30} years, so they are now pretty clearly excluded. But while violation of the law of baryon number conservation, and hence the possibility of proton decay, is a robust consequence of unification, the exact value of the predicted lifetime depends sensitively on details of how the basic idea is implemented in a concrete model. So we have the unsatisfactory situation that the theoretical idea of unification seems more attractive than ever, but there is no good idea for testing it experimentally.

How will we escape from this impasse? There are a few ideas for other experiments that could in principle reveal some sign of the physics associated with unification. Detection of cosmic axions or of neutrino masses, are genuine possibilities. These possibilities may already be physical realities as the mysterious cosmic "dark matter" could well consist of axions or massive neutrinos, and the absence of expected neutrinos from the sun may possibly indicate changes in their properties (oscillations) resulting from their having mass.

Also, some of the exotic objects associated with

unification—magnetic monopoles or cosmic strings—might have been produced in early states of the big bang and persisted to the present day. Experimental searches for such exotics are being rigorously pursued. On the other hand, one of the main advantages of the inflationary universe idea (discussed in Chapter 11) is precisely to get rid of these objects, which would otherwise excessively clutter up the universe.

Finally, it is possible that a unified theory will emerge that is so complete and compelling that its consequences, including violation of the law of baryon-number conservation, will be accepted even without direct tests. Such a theory would, of course, have to have other consequences that can be tested directly. It might, for example, predict the ratio of electron to muon mass, the observed magnitude of CP symmetry violation or the existence and properties of hitherto unobserved particles at future accelerators. At present many physicists attach high hopes to superstring theories, but as yet no concrete predictions have been extracted from these theories.

In summary, the attractive speculations discussed in my article remain attractive speculations. There is considerable hope, but no certainty, that within the present millenium they will become more than that.