SLAC–PUB–8784
August 2001

The Symmetry, or Lack of It, between Matter and Antimatter*

Helen R. Quinn
Stanford Linear Accelerator Center
Stanford University, Stanford, California 94309

Invited talk presented at the
15th Anniversary Meeting of Dirac Medalists
ICTP, Trieste, Italy
November 4–5, 2000

---

*Work supported by Department of Energy contract DE–AC03–76SF00515.

The subject of antimatter and its relationship to matter began with Dirac, with the publication of his famous equation in 1928.[1] Today it remains an active area of particle physics. The dominant issue for a number of major experimental programs is to decipher the nature of the difference in the laws of physics for matter and for antimatter. This has been a central issue of my work in the past few years, and a recurring theme in earlier work. Hence when I was asked to review a subject of my choice for this conference, this was the obvious choice for me; a very different focus from any other talk here. (Also, it allows me along the way make reference to both pieces of work for which I was cited in my Dirac award, though neither is central to this story.) Given this opportunity, I decided to start with the early history of the subject, both in honor of Dirac and his essential role in it, and because it is fascinating to look back and see how understanding evolves.

Today the closest thing we have to a symmetry between matter and antimatter is CP symmetry. It is not an exact symmetry but rather one that is broken, albeit only by small effects in the weak interaction sector (and by a possible tiny $\theta\epsilon_{\mu\nu\lambda\sigma}F^{\mu\nu}F^{\lambda\sigma}$ term that gives rise to strong CP violations).[2] The fact CP symmetry it is violated in weak processes was discovered by Christensen, Cronin, Fitch and Turlay in 1964.[3] The discovery shocked particle physics and opened the door to questions still at the core of particle physics and of cosmology today. Cronin, speaking of this discovery in his 1980 Nobel lecture said: *"we are hopeful ... that at some epoch, perhaps distant, this cryptic message from nature will be deciphered."* [4]The cryptic message is the lack of an exact CP symmetry, but also that fact that it is so nearly a symmetry. We live in a universe where matter clearly dominates over antimatter, so both the near symmetry and the nature of its breaking, in the laws of physics and in the history of the Universe, are issues of deep interest.

Prior to Dirac there was no concept of such a symmetry, or indeed of the existence of antimatter. Instead there was simply a conservation law, the conservation of matter (or of atoms). Thus all the matter, all the stuff with mass in the Universe, must always have been there. At the time in question the prevailing view of western science, and indeed of western religions and philosophy, was that we live in a static unchanging universe. So the constancy of matter in this Universe was no puzzle, simply a consistent piece of the picture. Hubble's law, the linear relationship of redshift of distant galaxies with the distance to them,[5] provided the first key evidence for an expanding Universe. It appeared a year after Dirac's equation. With this discovery the science of cosmology began; questions about the physical evolution of the Universe began to be asked. But even in an evolving Universe, if matter is conserved, all the matter present today must have been present at the beginning. Its presence can only be understood as an initial condition of the evolution.

To further set the context for Dirac's work and the struggle to interpret it, it helps to recognize that at this time there were only two known fundamental matter particles, electrons and protons. The nature of nuclei, and the concept of neutrons and their discovery, was a work in process, but not yet concluded. The neutron was only detected, and the detection understood, in 1932.[6] So at the time Dirac's equation appeared physicists did not readily postulate a new particle types to explain new phenomena, let alone to explain a peculiar result in a newly-postulated equation.

The Dirac equation was the fruit of Dirac's struggle to obtain a relativistic spin 1/2 equation of motion. The equation had one very successful feature, as Dirac later said *"an unexpected bonus, completely unexpected"* (Dirac reminiscing, 1977).[7] It gave the correct magnetic moment for an electron. This was strong evidence in its favor. But at the same time it contained an enigma, which, as you all know, manifests itself at the first level of understanding as the existence of negative energy states. Such states are clearly unphysical. At best they are an indication that one is attempting to do perturbation theory around the wrong "vacuum". At worst they say the theory is incurably sick. For a theory of fermions the exclusion principle offered a cure. A better vacuum or ground state can be found, in which all negative energy states are filled, but all positive energy states are empty. That is clearly the lowest energy state.

This does not completely remove the enigma. Instead it transforms it. Dirac's equation with this ground state has excitations that must be interpreted as positively charged particles (or holes in the negative-energy sea) in addition to the negatively-charged electron states. What possible physical interpretation could be made of these objects? As Dirac later stated *" At that time ... everyone felt pretty sure that the electrons and the protons were the only elementary particles in nature"* [8] The only choice, it seemed, was that the positively-charged states were protons. This interpretation was indeed the first one offered. It had one obvious deficiency, and a second, and more fatal, flaw that took a little longer to be noticed.

The first and most obvious problem was that the equation seemed to say that the proton and the electron have the same mass. Indeed it gave an exact symmetry between the properties of the positive and negatively charged states. Dirac was not unaware of this, as his 1929 letter to Bohr makes clear: *"as long as one neglects interactions one has complete symmetry between electrons and protons; . . . However when the interaction between the electrons is taken into account, this symmetry is spoilt. I have not yet worked out mathematically the consequences of the interaction .... One can hope, however, that a proper theory of this will enable one to calculate the ratio of the masses of protons and electrons."* [9] Here he expresses his hope that the symmetry in his equation can be removed by including a next-order correction—I can only guess that he means something like interactions between the particles filling the negative energy states. Whatever he might have been thinking his guess was completely wrong. The symmetry was there to stay.

The problem just would not go away. As Weyl stated it, in November of 1930, *"...indeed, ... the mass of the proton should be the same as the mass of the electron; furthermore ... this hypothesis leads to the essential equivalence of positive and negative electricity under all circumstances"* (in modern language exact matter-antimatter symmetry) *"... the (observed) dissimilarity of* *the two kinds of electricity thus seems to hide a secret of nature which lies deeper than the dissimilarity of past and future … a new crisis of quantum physics…"* [10]

The crisis deepened when Oppenheimer and, independently, Tamm noticed the second problem.[11, 12] They saw that the equation contained the possibility of particle-hole annihilation. This was the kiss of death for the proton interpretation of the holes. Even if corrections might fix the mass, they could not remove this disaster. This process, if it existed for proton and electron, would destroy all possibility of stable matter. Hydrogen atoms, or any others for that matter, could simply dissappear in a flash of light. That clearly does not happen. A new approach to the holes was required.

So eventually Dirac, in May 1931, made what he later called a *"small step forward"*. [13] He declared that *"A hole, if there were one, would be an entirely new kind of particle, unknown to experimental physics, having the same mass, and opposite charge of the electron."*[14] The evasive phrase "if there were one", signals, it seems to me, his ongoing resistance to making such a dramatic prediction. The idea of predicting a new type of particle was so overwhelmingly strange that he felt the need to soften his statement in this way. But it was not long before that evasion became unnecessary.

Not much more than a year later, these particles were observed. In a paper by Anderson, submitted Feb 1933; published in PRL 43, March 1933,[15] we find the evidence: *"On August 2, 1932 ldots tracks were obtained .. which seemed to be interpretable only on the basis … of a particle carrying a positive charge but having a mass of the same order of magnitude as that normally possessed by a free negative electron."* and indeed more than one event *"… other photographs were obtained which could be interpreted logically only on the positive-electron basis…"*. Anderson coined a new name for his discovery *"… the magnitude of the charge of the positive-electron which we shall henceforth contract to positron, is very probably equal to that of a free negative electron which from symmetry reasons would then naturally be called a negatron."* Anderson's renaming of the electron was ignored, but the name positron for its antiparticle is still used. This, the first known antiparticle, is the only one honored by having a distinct name, rather than just anti-electron.

The observation of the positron raised a new question. Why is the world populated with electrons but not with positrons? This became for the first instance of the cosmological puzzle of the dominance of matter over antimatter. This cosmological issue was recognized by Pauli, who found it unpalatable enough that he rejected the theory that had made the dramatic and correct prediction of these particles. He stated in a June 1933 letter to Heisenberg: *". . . I do not believe in the hole theory, since I would like to have the asymmetry between positive and negative electricity in the laws of nature (it does not satisfy me to shift the empirically established asymmetry to one of the initial state)."* [16] I find this a quite remarkable statement, particularly its parenthetical coda. As far as I know it is a first statement of the view, held today by many particle physicists and cosmologists, that it is unsatisfactory to appeal to initial conditions as an explanation for the dominance of matter over antimatter in the Universe. Pauli also points out another crucial point, that in order to avoid having to appeal to initial conditions one must somehow remove the symmetry between matter and antimatter that is a feature of Dirac's equation.

Pauli's comment aside, most physicists for the next twenty-some years not only accepted the existence of antimatter, but also the exact symmetry, C or Charge Conjugation, between the laws of physics for matter and antimatter. Perhaps because this symmetry was indeed unavoidable in electromagnetic theories, it was commonly assumed to be an inevitable, and indeed desirable, feature of all particle theories. In QED, the field-theoretic version of Dirac's equation, the automatic C symmetry of Dirac's equation persists. In addition it was recognized that all local field theories have a compound symmetry CPT, where P is parity or co-ordinate reflection and T is time-reversal. In QED, all three of these symmetries are separately conserved. It seems most physicists expected this pattern to be a general one.

I will now leap ahead in time to reach the point at which this assumption was first challenged, with the discovery that weak interactions violate both the charge-conjugation symmetry C between particles and antiparticles and at the same time also violate P or parity. A few highlights from the development of physics over the intervening twenty-three years are needed.

On the experimental front much was learned. The development of particle physics research at accelerator laboratories, starting with the 350 MeV cyclotron at Berkeley, and of detectors for particles, starting from emulsions and cloud chambers for studies of cosmic ray secondary particles, opened the door to a multitude of new particle discoveries. The initial aim of the work was to understand nuclear forces. Discoveries of more and more particles pushed that goal to the side as the patterns and properties of the particles themselves became a field of study in its own right. This was not a distraction, it was a necessary detour. Only once the plethora of particles had been categorized, and their underlying structure deciphered could the modern theory of interactions be developed. With it, some answers to the original puzzle of nuclear forces could emerge.

Meanwhile there were further important developments concerning antimatter. The discovery of antiprotons and antineutrons removed any lingering doubt that the nucleons were, at least approximately, also described by Dirac's equation. Chamberlain, Segre, Weigand and Ypsilantis; (Oct 1955) *"The extension of the Dirac theory to the proton requires the existence of an antiproton"* although *" the anomalous magnetic moment of the proton indicates that the Dirac theory does not give a complete description of the proton."* But the predicted antiparticle was observed *"... based on a determination of the mass of negative particles ... simultaneous measurement of momentum and velocity... to date sixty antiprotons have been detected."* [17] Antineutrons were seen by Cork, Lambertson, Piccioni, Wenzel (Oct 1956) *"... the principle of invariance under charge conjugation gained strong support when it was found that the BeVatron produces antiprotons. Another prediction of the same theory ... (is) the existence of antineutrons."* The experiment was to *"detect the annihilation of antineutrons produced by charge exchange from antiprotons."* ($p + \bar{p} \rightarrow n + \bar{n}$).[18]

The form-factors and anomalous magnetic moments of the nucleons were problems to be understood much later, once their composite nature was elucidated. But these discoveries established the concept of antimatter particles as a general one, applicable for all fermions, not just electrons. This truly broadened the scope of the issue that concerned Pauli, the cosmological asymmetry between matter and antimatter species populations in the universe.

Notice that the experimental discovery of the antineutron was presented as a verification of matter-antimatter symmetry, viewed at that time as Charge conjugation or C invariance. The very existence of antimatter was taken to be evidence for the symmetry. After all it was this symmetry in Dirac's equation, that predicted the existence of the additional particle types. These discoveries also changed the basic law of conservation of matter. Annihilation and production of matter can occur, but only in with annihilation or production of a matching amount of antimatter. So the law of conservation of matter is corrected, but survives as the law of conservation of baryon and lepton numbers, the number of particles minus antiparticles of a given fermion type. These two features, symmetry between matter and antimatter in the equations, and baryon number conservation were considered exact laws. The observed asymmetry of matter and antimatter in the Universe could still only be accounted for by imposing an initial condition.

Among the new particles discovered (starting with the pion in 1947)[19] were the class known as mesons, integer-spin, massive particles. These could not be classified either as matter or as antimatter. However, one could consider the relationships among these particles under the operation of charge conjugation, C, and the language particle and antiparticle was extended to these situations. The charged mesons had particle-antiparticle pairings. Most neutral mesons, it was assumed, were self-conjugate under C. Like the photon, they each were their own antiparticles. However, among the neutral mesons one particularly peculiar pair, as physicists chose to call them the "strange" mesons, were the tau and theta. These shared the peculiarity of the charged strange mesons, K^+ and K^- by decaying relatively slowly, compared to other mesons of similar mass. The neutral tau and theta mesons had an added peculiarity, they were two apparently distinct states with different decay lifetimes. But the two particles had essentially equal masses and produced in the same ways.

Both puzzles were resolved by Pais and Gell-Mann in 1952–1956 by introducing a new quantum number, which eventually became known as strangeness.[20] This new quantum number is conserved in the strong interaction production processes but violated in weak decays. As far as particle-antiparticle relationships were concerned, any such additional quantum number led to additional pairings, as this quantum number, like electric charge, was a quantity that reversed sign under C. Then there could be neutral bosons that were not identical to their antiparticle. Thus the introduction of the strangeness quantum number explained the existence of two near-degenerate neutral states. These could be understood as two states of opposite strangeness: K^0 and its antiparticle $\bar{K}^0$. The two definite-strangeness states could be combined in two quantum superpositions that transform into themselves under C (and CP) transformations. These are the linear combinations $[K^0 \pm \bar{K}^0]/\sqrt{2}$.

Interesting quantum physics arises from the interplay of production of strangeness eigenstates, but propagation and decay of eigenstates of definite CP. This could explain the observations up till that time.

- states of definite strangeness are strongly produced but only as pairs of particles with equal and opposite strangeness
- states of definite CP have distinct masses and lifetimes
- only the CP-even state can decay to two pions, by conservation of CP.
- the phase space for three pion decay is limited

- thus the half-life of CP even state is much shorter than that of the CP-odd state.

This resolved the old $\tau - \theta$ puzzle, and, until 1964, appeared to describe the data well. In particular the two very-different half-lives, $\tau(K_{\text{Short}}) = 0.9 \times 10^{-10}s$ and $\tau(K_{\text{Long}}) = 5 \times 10^{-8}s$ could be explained by the assumed CP symmetry. (I use the modern language of CP symmetry here, although at the time it was first understood the assumed symmetry that "explained" all this was C.)

The long-half life of the charged strange particles was also explained by the strangeness quantum number. They have only weak decays, and suppressed ones at that. Cabibbo later pointed out that there was a triangular relationship between the coupling strengths of these two types of hadronic weak decays, strangeness-conserving (*e.g.* pion decay, which I denote as a) and strangeness-changing (K decay, b) and the weak decays of muons (c) of the form $a^2 + b^2 = c^2$.[21] This pattern led to the idea of a unitary mixing matrix which describes how the down quark states which pair definite up quarks under weak interactions are related to the down quark states of definite mass, a structure that became the basis of the Standard Model.

It is worth a comment that the discovery of "antimatter" and of mesons was a complete revision of the concept of "matter." In modern parlance some stuff with mass is matter—baryons and leptons, some is antimatter—antibaryons and antileptons, and then there is some other stuff, mesons, that have mass but are apparently neither matter nor antimatter! Clearly it is no longer satisfactory to define matter as that which has mass. By inventing the word "antimatter" particle physicists forever changed the definition of the word matter. Unfortunately, at least in the US, the definition of matter you find in most middle school text books is the out-dated one "matter is that which has mass and occupies space." Today, seventy-some years after the discovery of positrons, most people in the US still think antimatter exists only in the minds of science fiction writers. I find that many educated people are truly surprised to hear that it is real, and commonly made in our laboratories.

But what about the symmetry between matter and antimatter? As time went on, it was noticed that the fact that CPT is an exact symmetry in all field theories does not require that any of the three individual sub-symmetries is also exact. Lee and Yang further pointed out that there was indeed no evidence for or against parity conservation in weak interactions,[22] and that the assumption could be checked. The experiments followed soon thereafter. Wu, Ambler, Hayward, Hoppes and Hudson (January 15,1957) [23]published that "in . . . beta decays of polarized nuclei . . . asymmetry in the distribution . . . (gives) unequivocal proof that is not conserved in beta decay . . . this effect has been observed . . ." In the same issue of Physical Review we find a second paper by Garwin, Lederman and Weinreich (January 15, 1957)[24] which states that they have made "Observations of the failure of parity and charge conjugation . . . the magnetic moment of the free muon."

While P and C fail in these experiments, it was recognized immediately that CP survives; the weak interactions involve only left-handed neutrinos and only right-handed antineutrinos. Thus, despite the failure of the original matter-antimatter symmetry, C, there is a new matter-antimatter symmetry, CP, that appears to be exact. Physicists at the time were very happy with this result. Today one wonders why, once the P and C symmetry were found to be broken, the entire physics community did not expect that CP also would eventually fail; it seems physicists are slow learners! I am no historian, but my guess is that the loss of P symmetry was so great a surprise that physicists were happy not to face the further question of CP breaking until later. For one thing CP symmetry maintained the nice explanation of the neutral kaon lifetimes. Furthermore CP symmetry breaking would also imply T-symmetry breaking, and somehow that seemed unattractive. One could find a theory, the four-Fermi theory of weak interactions, that fit the observations and preserved CP symmetry and T symmetry. As far as I know no-one immediately made the leap that Pauli had earlier made, to the idea that a theory without CP symmetry would be preferable. Not until Sakharov's 1967 paper (discussed below), which comes after the empirical discovery of CP violation do we find this cosmological question addressed again.

The comfortable assumption that CP is an exact symmetry of nature was destroyed in 1964. CP-symmetry violation was discovered in neutral K meson decays. The experiment of Christenson, Cronin, Fitch and Turlay, PRD July 1964 [3] searched for the CP-forbidden two-pion decays of K_{Long}. A beam of mesons begins as an equal mixture of K_{Short} and K_{Long}. By studying the beam a sufficiently long time after its production (sufficiently far from the production point) one can reduce the K_{Short} content to as small a fraction as desired. But two-body decays were seen in a proportion significantly higher than the residual K_{Short} fraction. The authors *"... conclude therefore that* K_2 *[*K_{Long}*] decays into two pions with a branching ratio ...* $2.0 \pm 0.4 \times 10^{-3}$*."* Further they state *"The presence of a two-pion decay mode implies that the* K_2 *meson is not a pure eigenstate of CP."*

But if the mass eigenstate is not a CP eigenstate then the only possible conclusion is that CP symmetry is not exact. The laws of physics have no exact symmetry between matter and antimatter! At the time, this was an astounding result. It is easy to say today that physicists could well have anticipated it, once P and C symmetry violations had been seen, but no-one had done so. All particle theories at the time had natural CP symmetry, with no obvious freedom to add any CP-violating parameters. The measurement that found this result began as an effort to lower the experimental upper limit on the rate for this decay, but instead found that it does indeed occur. The experimental result was simple and irrefutable, and rapidly confirmed by others. No matter what the theoretical prejudice might be, the effect was real. The modern theory of particle interactions, the three-generation Standard Model, can readily accommodate this effect, but that was not evident till almost ten years later.

Once this effect had been seen, the possibility that the matter-antimatter asymmetry of the Universe was a result of its evolution rather than simply the initial condition so disliked by Pauli could be realized. One of the first people to recognize this was Sakharov. In 1967 he proposed that baryons and antibaryons were present in equal quantities in the early universe, and that the imbalance developed at some later time, a process nowadays referred to as "baryogenesis". Sakharov[25] showed that this requires

- baryon number changing processes in the early Universe
- CP Symmetry Violation
- an out of equilibrium situation for the universe at time of baryogenesis

Sakharov's primary observation was that it follows from the equal masses of particles and their antiparticles that, if the two species are in thermal equilibrium, their populations are equal. If there is any baryon number-changing process then this thermal equilibrium will be achieved. This means that, whatever the initial condition, collisions and decays drive the baryon number (and, locally, the baryon number density) to zero. One must protect an initial condition with a conservation law for it to hold sway throughout the history of the universe.

Sakharov further recognized that any transition from thermal equality of matter and antimatter to the present inequality must occur at an out-of-equilibrium stage in the history of the universe, or else the equilibrium condition of equality is maintained. Further, for any imbalance to develop, both baryon-number conservation and CP symmetry must be broken. CP symmetry would require a balance between any baryon-producing process and its CP inverse, antibaryon-production. Finally, baryon number-changing processes must be rare, or "frozen out" at any time after the transition time. If not, these processes will gradually remove any baryon excess and return the system to the naive thermal equilibrium between the baryons and antibaryons, namely zero baryon number. For its time, this was a revolutionary paper, since the notion of baryon-number conservation was still firmly established in the theorist's canon. The stability of matter, with a proton half-life of over 10^{30} years, seemed to justify that particular prejudice.

Now to discuss whether, and if so when, Sakharov's three conditions are satisfied, one must have a fully-developed theory of particle interactions that includes both CP-symmetry breaking and baryon-number changing terms. This we have, the Standard Model of particle physics. This theory, which evolved over the years from 1964 to 1973, has been spectacularly successful in providing the basis for interpreting all experimental results. This is not the place for a full history of the ideas and experiments that culminated in this theory; instead I focus on the issue of matter-antimatter asymmetry as it manifests itself in this context. The early version of the theory was a two-generation theory, with four quarks and four leptons. Of these the fourth quark was a putative particle, added to the theory to avoid the problem of strangeness-changing neutral currents. The discovery of particles containing this quark in 1974 [26] was the first major success of the modern Standard Model, although it took over a year of work to be sure that this interpretation of the J/ψ particle was the correct one.

But in fact, even in 1973, the two-generation Standard Model could not have been the full theory, because, it turns out, this theory, like Dirac's equation, automatically has an exact CP symmetry. Most physicists at the time ignored this problem, there was so much else to investigate about this proposed theory. Others remained skeptical of quark-based theory, because of the problem of fractional quark charges. (The quark-parton interpretation of deep inelastic scattering experiments, which we now clearly see as indicating quark structure in protons and neutrons, was not fully developed until 1969, and not widely appreciated until some time after that.) The proposed fourth quark seemed to many to be a highly speculative idea.

Kobayashi and Maskawa, however, did address the issue of CP symmetry breaking in a little-noticed paper in 1973.[27] First they pointed out that the theory with two generations was automatically CP-conserving, then they explained that a three-generation generalization of the Cabibbo structure of quark weak currents gave a theory that allowed CP violation, but with only a single CP-violating quantity. This at a time when most physicists were skeptical of the hypothesis of a fourth quark. But prior to the discovery of charm little attention was paid to this work. Weinberg later observed that another option, instead of adding additional quarks, was to add additional Higgs bosons.[28]

The big change in attitude that occured in 1974–1975 was not just due to the discovery of particles containing charm. Sorting out the rate of hadronic events in e^+e^- collisions above the threshold for production of charm particles took over a year because of the remarkable coincidence that another new particle, the tau lepton, was being pair-produced in this same energy region.[29] Only once the tau-produced events had been recognized and removed could the expected charm-decay patterns emerge! So by the time the second generation of particles was complete we had also found a third lepton. The idea of proposing a third generation for the Standard Model was no longer a fringe suggestion. Kobayashi and Maskawa's three generation theory became the Standard Model, and with it, CP violation became an effect that could be included in theoretical discussions.

The known effect in K-decays is certainly consistent with this theory. The weak interaction physics is calculable and at the quark level the theory is entirely predictive. The problem is that we don't see quarks, what we observe are hadrons. This means that strong interaction physics enters the picture. We must relate our quark calculation to a hadron observation. Often this relationship contains effects which we cannot calculate in perturbation theory, the long-range effects of strong interaction physics. At the quark level there are one loop diagrams that convert a K^0 meson with quark flavor contents $\bar{s}d$ into a $\bar{K}^0$ meson with quark flavor content $\bar{d}s$ via exchange of two W mesons. This can be represented at the hadron level as a calculable coefficient times the matrix element of a local four-quark operator between a K^0 and a $\bar{K}^0$ state. But the matrix element is not readily calculable, so the connection between the measured mass differences for the two kaon eigenstates and the parameters of the W-couplings in the underlying quark-level Lagrangian is not simple. It has what are called "hadronic uncertainties", in this case the uncertainties in the calculation of the hadronic matrix element of a well-defined quark operator.

The calculation of such quantities has improved steadily over the years, with lattice techniques giving us today the most reliable answers. But uncertainties from hadronic effects continue to limit our ability to compare different measurements that depend on the same Standard Model parameters. Is the single CP-violating parameter in the Kobayashi-Maskawa scheme the full story, or are other extensions of the theory, such as Weinberg's additional Higgs particle, needed? This is very much an active question in current work, both theory and experiment. Within the ranges given by the hadronic uncertainties we can still fit all results with a single set of Standard Model parameters, but both new measurements and new calculations continue to refine the issue.

As an aside here I cannot resist a comment on the issue of strong CP-violation, partly because that allows me to include here a piece of work for which I was cited in my Dirac award. Once CP is not a symmetry of the full Lagrangian one cannot, in general, protect a gauge theory such as QCD from developing an additional $\theta\epsilon_{\mu\nu\lambda\sigma}F^{\mu\nu}F^{\lambda\sigma}$ term in the effective Lagrangian. Experimentally the upper bound on an electric dipole moment for the neutron places severe restrictions on the existence of any such term. Thus the smallness of this effect becomes a puzzle to be solved, and its solution is still an open question, for a review see [30]. One class of solutions, suggested by me and Roberto Peccei, involve an additional approximate U(1) symmetry and a concomitant light and very weakly-interacting particle, the axion.[31] Other approaches postulate that CP is an exact symmetry of the theory that is spontaneously broken by soft-breaking terms only. This can result in a sufficiently small induced theta term. The size of this term depends on details of the theory. The third possible solution is that the bare mass of the up quark is actually zero, which then makes the theta term effectively zero; the observed up quark mass is then due to corrections induced via weak interactions. This is numerically disfavored by the pattern of best-fit quark masses, but not completely ruled out.[30]This is another place where the answer awaits further experimental results, constraints on axions from experiment and from astrophysical bounds have left only a narrow window; and axion search experiments are beginning to reach interesting sensitivity.

Once we have a theory that includes CP violation we can address the question of the matter-antimatter asymmetry in the Universe. CP violation was only one of Sakharov's conditions for generation of this asymmetry. Equally striking, and in some ways more significant, was the requirement that, baryon number not be a conserved quantity. With the introduction in the early seventies of the idea of Grand Unified Theories, this idea gained currency. In most grand unified theories (GUTs) the proton is unstable, but has a very long half-life due to the high mass of the additional gauge bosons of the GUT, compared to those in the Standard Model sectors. Indeed the advent of these theories which merge all three gauge interaction types into a single gauge group was a significant step in particle physics thinking and prompted renewed attention to searches for proton decay. The calculation of the merging of the coupling constants at very high scale (recognized in the Dirac awards for Howard Georgi and for me) provided an estimate of the mass of the new bosons and thus of the proton half-life in any such theory.[32] Measurements have now pushed the proton half-life limit to greater than 10^{32} years, which rules out the simplest (non-Supersymmetric SU(5)) realization of the Grand Unified Theory idea, but many possible variants survive. There is still no direct evidence that the Grand Unification idea is correct, but it is so attractive an idea that it cannot be ignored.

Despite their rarity today, at sufficiently high temperature in the early Universe, such baryon-number-changing processes would be frequent. Then thermal equilibrium with equal amounts of matter and antimatter would prevail at high temperature in the Universe, no matter what the starting condition. So the existence of baryon-number-changing processes which Sakharov showed is necessary to explain how an imbalance could develop, also makes that explanation essential, unless an initial condition was protected by some other conservation law.

The third condition, that of an out-of-equilibrium situation, can be met, for example, whenever the Universe undergoes a phase transition. When modern particle physics theories are examined in a cosmological context there are a number of such transitions. These occur as the universe expands and thus cools. In particular the various stages of symmetry breaking in a GUT theory occur as phase transitions, as does the transition from a quark-gluon plasma to gas of hadrons (actually still a charged plasma from the electromagnetic point of view, but a color-singlet and therefore neutral gas with respect to QCD charge.) The focus for understanding the matter-antimatter asymmetry was, at first, on the earliest of these phase transitions, since massive-boson-mediated baryon-number- changing processes are frozen out once these baryons become massive compared to the prevailing temperature ($M > kT$).

It was not long however before it was recognized that even well below this temperature there are surviving non-perturbative baryon-number changing effects. Instanton-like configurations (sphalerons) of the Standard Model gauge fields can change baryon number. While very unlikely at low temperature these configurations are not suppressed at high temperature. So even in the Standard Model, at high enough temperature, baryon number is not conserved. These non-perturbative effects are not frozen out until after the weak phase transition, in which the W and Z as well as all quarks and leptons become massive. Calculations showed that even if a baryon excess was generated at high temperature these non-perturbative effects would steadily reduce it, back towards the zero value of thermal equilibrium. The eventual result was much too small an asymmetry. Thus efforts to explain the baryon excess were refocused on late stage phase transitions, such as the weak transition.

The typical weak phase transition baryogenesis scenario depends on this being first order phase transition. At high enough temperature the vacuum has a vanishing value for the Higgs field, whereas at low temperature the preferred vacuum value is non-zero. In a first order phase transition a bubble region forms with the correct low-temperature vacuum inside it, and a distinct bubble wall. Outside the bubble is a region where baryon-number-changing processes are not suppressed. Inside, because the W and Z have large mass, these processes are essentially non-existent, so any non-zero baryon number that develops inside the bubble will be maintained. Thus inside the bubble there is no simple thermal equilibrium between baryons and antibaryons, instead there is conservation of baryon number. As the bubble expands its surface sweeps through space. The imbalance in matter over antimatter inside the bubble arises because CP violation gives a differential transmission probability for baryons vs antibaryons through the bubble wall. Eventually the surface recedes beyond our horizon, so our entire observable universe is inside this bubble.

Within the Standard Model one can calculate (roughly) what imbalance would arise.[33] The number to be calculated is the ratio of entropy in matter to that in radiation, radiation that we now observe in the form of the 2.7K background microwave radiation. The energy of the Universe is today dominated by matter over radiation. This is because with expansion radiation redshifts, but mass does not. However, the ratio of baryons to entropy in radiation is an extremely small number, indicating an initial excess of about 1 baryon in 10^{10} baryon-antibaryon pairs. Even so, when the experimental constraints on the Higgs mass are included, the scenario described above fails to generate a large enough baryon number.[34]

In fact the most recent limits on Higgs mass go further; they destroy the entire scenario. With this constraint one finds that the relevant phase transition is probably not first order. The picture of an expanding bubble with a well-defined bubble wall is called into question. In the minimal Standard Model the scenario simply does not work. For particle physicists this suggests a fascinating possibility—perhaps the failure is not in the cosmological scenario but in the underlying particle physics of the Standard Model. For example if one adds additional Higgs field then the constraints on the

Higgs mass and hence on the phase transition are changed. Furthermore this provides additional parameters that can introduce CP violation into the theory, beyond the one such parameter in the CKM matrix. Perhaps one needs Weinberg's idea for a source of CP violation in addition to Kobayashi and Maskawa's. Of course one can also investigate even more complex extension of the Standard Model, among them supersymmetric theories. The list of ideas is long; and only further particle physics experimentation can narrow it down.

So this brings us back to the fundamental particle physics question, what is the asymmetry in the underlying laws of physics between matter and antimatter? Is it correctly and fully described by the CP-violation in the CKM matrix in the Standard Model or not? This is a question whose answer we can explore in the high energy laboratory –or at least some part of the answer can be sought there. Indeed this is the chief aim of the large effort focused on B physics in high energy laboratories around the world. The aim of this effort is to make as many redundant measurements as possible of quantities that, in the Standard Model, determine the values of CKM matrix elements, including the one CP-violating phase, (or the equivalent rephasing-invariant quantity known as the Jarlskog invariant).[35] The aim is to find whether all these measurements give results consistent with a single set of values for the four independent parameters that define the CKM matrix. If so we will have better-determined values for all these quantities. If not, then further study of the nature of the inconsistencies will point toward one or another possible addition to the Standard Model theory—and thus perhaps offer a better way to understand Baryogenesis too.

This work is well begun, the SLAC and KEK B factories have produced first results[36] and will yield additional interesting results over the next five to ten years. These will be complemented by other results from B-physics experiments at hadron colliders, such as the mass and width differences of the two B_s mesons, mesons which are not produced by the e^+e^- colliders at their current energies. B physics provides a wonderful laboratory to study CP violation because both the B_d and B_s mesons form a pair like the neutral K mesons where the flavor eigenstates are mixed to form mass eigenstates, and the differences of the mass eigenstates from exact CP eigenstates provides a sensitive probe for CP violation. Further, because of the large mass of the b quark, the B mesons have many possible decay channels so they provide many different probes of the Standard Model CKM and CP-violation structure. This allows one to develop the redundant set of independent determinations of parameters that are needed to probe for consistency. The impact of long-range hadronic physics, which confounded this effort in the kaon case, is not totally removed. However, because of the larger B mass its impact better controlled than in the Kaon case, another reason why particle physicists find studies of B decays of such compelling interest.

An intriguing new possibility has recently appeared. Three different types of evidence now all point to the fact that neutrinos have some tiny masses, and that, as for the quarks, the mass eigenstates and the weak-interaction-decay eigenstates are misaligned. This phenomenon is usually referred to as neutrino-oscillation. Neutrinos are produced as particles with definite flavor but appear to oscillate to some other flavor and back as they travel through space (or through matter) due to the coherent propagation of the different mass eigenstates.[37] The masses are too small to be measured directly, but observations can tell us something about the mass differences between two mass eigenstates that both contain some component of the original flavor eigenstate, and about the mixing.

It is worth remarking that the language for the neutrino masses is a little different from the quark case, chiefly for historical reasons. Our particle names are based on the ways in which these particles are produced. For the quarks we assign names to the mass eigenstates and treat the weak-interaction eigenstates as mixtures of these. We do this because we see quarks produced as flavor eigenstates via strong interactions. For the neutrinos we have traditionally named the distinct states by the flavor of their charged-lepton weak-interaction partner. Indeed this is the only way we could distinguish them if they were massless particles. Since neutrinos are only produced by weak interactions, these are the states we observe at production. Now that it appears neutrinos have some small mass, we seek to determine what mixtures of weak-eigenstates form the definite-mass eigenstates. The two situations are mathematically very similar. The complete description of the system is given by the set of eigenmasses plus a three-by-three matrix that tells us the transformation between the weak-interaction eigenstates and the mass eigenstates. This is the CKM matrix for the quarks; a similar matrix exists for the neutrinos. Just as the Standard Model with zero neutrino masses had possible CP violation in the CKM mixing-matrix structure, so too the neutrino mixing matrix may contain additional CP violation.

The evidence for neutrino masses and mixing comes from three distinct types of experiment. The flux of neutrinos from the sun is roughly half that predicted from models. This result is now observed essentially over the entire range of neutrino energies. It is thus very hard to accommodate by any adjustment of solar model parameters.[38] However it can be easily understood if electron neutrinos oscillate into some other type as they travel through the dense matter of the sun, or through the space between sun and earth. The second result has to do with neutrinos produced as tertiary particles from the decay of the mesons produced in the collisions of high energy cosmic rays with the upper atmosphere. The ratio of electron-type to muon type neutrinos, and the dependence of this ratio on azimuthal angle can be modeled with and without neutrino-mixing oscillation effects. The evidence suggests that muon neutrinos are oscillating away, but not producing additional electron neutrinos. The third result, from reactor-produced neutrinos, also suggests an oscillation in which electron-type neutrinos are disappearing and some other type produced. The overall situation at present is a little confusing. The three experiments give three different scales of mass difference, so they cannot be consistently interpreted with three neutrino species (and thus only two independent mass differences).[37] But be that as it may, the evidence that there is some non-zero mass and mixing matrix is steadily growing. A long-term program of experiments will be needed to determine as precisely as possible all the parameters of the neutrino masses and mixings. In particular it will be very difficult to develop sufficient sensitivity to measure the CP-violating effects from this mixing matrix. But the possibility that there is such an effect opens the door to new scenarios to explain the universal dominance of matter over antimatter.

Neutrino masses provide a very different scenario to fulfill Sakharov conditions. In GUTS that contain right-handed neutrinos these particles get large Majorana-type mass. Then the small masses of the weakly-interacting species can be understood. The small left-right coupling from the Higgs and other scalars generates a small mixing between the heavy neutrinos and the massless left-handed ones. This gives one heavy and one very light eigenstate per flavor. Early enough in the evolution of the Universe the massive neutrino species are in equilibrium with all other matter. However, since they have only a tiny component of the weakly-interacting left-handed particle they interact very weakly. Hence they "freeze out", that is drop out of equilibrium, very early on. Then the CP-violation in their decay produces a lepton-anti-lepton asymmetry and a net lepton number in the remaining light species. At this stage, all other $(B-L)$-changing interactions have been frozen out (have become improbable) so this quantity is now fixed. The lepton asymmetry then thermalizes via processes that change both baryon and lepton number, but not their difference, for example the non-perturbative sphaleron processes. This then produces a baryon excess. This interesting possibility that been explored in a number of GUT scenarios. The details depend on the particular realization of a Grand Unified model and its parameters. It is possible to find scenarios that give the observed asymmetry.[39] Time will tell if these models are satisfactory in other aspects of their predictions as well.

To summarize: CP Violation is now a well-established effect. The laws of physics for matter and antimatter differ slightly. The simplest Standard Model theory of particle physics accommodates this effect with just one parameter. We do not yet know if this is the correct picture for CP violation. Experiments underway in B physics are one arena that can help probe that question. Neutrino masses and mixing pose another arena for CP violation, and another long-range program of experiments will be needed to explore this aspect of particle theory. It is very likely that the early universe had matter-antimatter equality, enforced by thermal equilibrium in any theory with any baryon-number changing processes. Then the currently observed asymmetry had to be generated, either by baryogenesis or by leptogenesis. We don't yet have a fully convincing story as to how or even when this occured. It is certainly a mystery worth solving!

Acknowledgments

In preparing this lecture I have been particularly indebted to the book "Paul Dirac, the Man and His Work" (Cambridge University Press 1998) and to the lecture by Abraham Pais therein. Bram's scholarship on this subject has been invaluable to this non-historian.

References

[1] P.A.M. Dirac, Proc. of Royal Society London, **A117**, 610 (1928); **A118**, 351 (1928).

[2] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976) and Phys. Rev. **D14**, 3432 (1976).

[3] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, Phys. Rev. Lett. **13**, 138 (1964).

[4] J. W. Cronin, Rev. Mod. Phys. **53**, 373 (1981).

[5] E. Hubble, Proc. Nat. Acad. Sci. **15**, 168 (1929).

[6] J. Chadwick, Nature **129**, 312 1932; Proc . Royal Soc. London **A136**, 428 (1932).

[7] P.A.M. Dirac, Report KFKI-1977-62 Hung. Acad. Sci.

[8] P.A.M. Dirac in History of Twentieth Century Physics, Academic Press (1977) p144.

[9] P.A.M. Dirac, letter to Bohr, November 26, 1929, copy in Niels Bohr library.

[10] H. Weyl, "Theory of Groups and Quantum Mechanics", p263-64 and preface Dover NY (1931).

[11] J. R. Oppenheimer, Phys. Rev. **35**, 562 (1930).

[12] I. Tamm, Zeitschr. fur Phyz. **62**, 545 (1930).

[13] P.A¿M. Dirac in interview with A. Pais, May 1964 as quoted by Pais in "Dirac, The Man and his Work", Cambridge University Press (1998).

[14] P.A.M. Dirac, Proc. Royal Soc. Lond **A133**, 60 (1931).

[15] C. Anderson, Science **76**, 238 (1932) and Phys. Rev. **43**, 49 (1933).

[16] W. Pauli, letter to W. Heisenberg June 16, 1933, Wolfgang Pauli, Scientific Correspondence, Springer (1979) Vol. 2, p169.

[17] O. Chamberlain, E. Segre, C. Wiegand and T. Ypsilantis, Phys. Rev. **100**, 947 (1955).

[18] B. Cork, G. R. Lambertson, O. Piccioni and W. A. Wenzel, Phys. Rev. **104**, 1193 (1957).

[19] C.M.G. Lattes *et al.*, Nature **159**, 694 (1947) and Nature **160**, 453, 486 (1947).

[20] A. Pais, Phys. Rev **86**, 663 (1952); M. Gell-Mann, Phys. Rev. **92**, 833 (1953); M. Gell-Mann and A. Pais, Phys. Rev. **97**, 1387 (1955).

[21] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[22] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956).

[23] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).

[24] R. L. Garwin, L. M. Lederman and M. Weinrich, Phys. Rev. **105**, 1415 (1957).

[25] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967).

[26] J. J. Aubert *et al.*, Phys. Rev. Lett. **33**, 1404 (1974). J. E. Augustin *et al.*, Phys. Rev. Lett. **33**, 1406 (1974).

[27] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[28] S. Weinberg, Phys. Rev. Lett. **37**, 657 (1976).

[29] A. Barbaro-Galtieri *et al.*, Phys. Rev. Lett. **39**, 1058 (1977). M. L. Perl *et al.*, Phys. Lett. **B70**, 487 (1977).

[30] For a recent review see M. Dine, hep-ph/0011376.

[31] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[32] H. Georgi, H. R. Quinn and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).

[33] For an introductory lecture on baryogenesis, see A. G. Cohen, http://www.slac.stanford.edu/gen/meeting/ssi/1999/media/cohen.pdf

[34] See for example M. Dine, R. G. Leigh, P. Huet, A. Linde and D. Linde, Phys. Rev. D **46**, 550 (1992) [hep-ph/9203203].

[35] C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985).

[36] A. Abashian *et al.* [BELLE Collaboration], hep-ex/0102018 (2001). B. Aubert *et al.* [BABAR Collaboration], hep-ex/0102030 (2001).

[37] For a recent review of these points and references to the data see M. C. Gonzalez-Garcia, hep-ph/0010136.

[38] J. N. Bahcall and P. I. Krastev, Phys. Rev. D **53**, 4211 (1996). For a more recent review of solar modeling see J. N. Bahcall, M. H. Pinsonneault and S. Basu, astro-ph/0010346.

[39] For a review see W. Buchmuller and M. Plumacher, hep-ph/0007176 (2000).