The Discovery of Quarks

Michael Riordan

Quarks are widely recognized today as being among the elementary particles of which matter is composed. The key evidence for their existence came from a series of inelastic electron-nucleon scattering experiments conducted between 1967 and 1973 at the Stanford Linear Accelerator Center. Other theoretical and experimental advances of the 1970s confirmed this discovery, leading to the present standard model of elementary particle physics.

Just 20 years ago, physicists were beginning to realize that the protons and neutrons at the heart of the atomic nucleus are not elementary particles after all. Instead, they appeared to be composed of curious pointlike objects called "quarks," a name borrowed from a line in James Joyce's novel *Finnegans Wake*. First proposed in 1964 by Gell-Mann (*1*) and Zweig (*2*), these particles had to have electrical charges equal to 1/3 or 2/3 of that of an electron or proton. Extensive searches for particles with such fractional charge were made during the rest of the decade—in ordinary matter, in cosmic rays, and at high-energy accelerators, all without success (*3*). But surprise results from a series of electron-scattering experiments, performed from 1967 through 1973 by scientists from the Massachusetts Institute of Technology (MIT) and the Stanford Linear Accelerator Center (SLAC), began to give direct evidence for the existence of quarks as real, physical entities (*4*). For their crucial contributions as leaders of these experiments, which fundamentally altered physicists' conception of matter, Jerome Friedman and Henry Kendall of MIT and Richard Taylor of SLAC were awarded the 1990 Nobel Prize in Physics.

The Prediction of Quarks

By the beginning of the 1960s, physicists had shown that protons and neutrons (known collectively as "nucleons") had a finite size of about 10^{-13} cm, as indicated by elastic electron-nucleon scattering experiments of Hofstadter and his Stanford co-workers (*5*), but the great majority considered these particles to be "soft" objects with only a diffuse internal structure. Along with pions, kaons, and a host of other "hadrons" (particles that feel the effects of the strong nuclear force), they were thought by many to be all equally fundamental—composed of one another in what had been dubbed the "bootstrap model" of strongly interacting particles (*6*). Theories that tried to explain the growing variety of hadrons as combinations of a small set of fundamental entities were a definite minority until the MIT-SLAC experiments occurred.

In 1961 Gell-Mann and Ne'eman introduced a scheme known as SU3 symmetry (*7*) that allowed them to impose a measure of order on the burgeoning zoo of hadrons. In this scheme, particles with the same spin are grouped together, as if they are just the various distinct states of one and the same entity—similar to the way the proton and neutron can be regarded as merely two different states of the nucleon. Particles with spin-0, such as the pions and kaons, form a group of eight "mesons" called an octet, as do another group with spin-1 (that is, those with internal angular momentum equal to Planck's constant h divided by 2π); the proton and neutron are the lightest members of an octet of "baryons" with spin-1/2, and there is a group of ten spin-3/2 baryons known as a decimet. In effect, Gell-Mann and Ne'eman did for physics what Mendeleev had done for chemistry—invent a "periodic table" of the hadrons. Using this approach, they even predicted new particles that were later discovered with appropriate properties, buttressing the faith of the physics community in SU3 symmetry as a correct representation of physical reality.

In seeking a deeper explanation for the regularities of their SU3 classification scheme, Gell-Mann and Zweig invented quarks (*1, 2*). In this approach there are three fundamental quarks—dubbed "up" or u, "down" or d, and "strange" or s—and their antiparticles, the antiquarks. Mesons are built from a quark plus an antiquark, and baryons are composed of three quarks. The proton is a combination of two up quarks plus a down quark (written *uud*), for example, whereas the neutron is made of an up quark plus two downs (*udd*). By assigning a charge to the up quark of $+(2/3)e$ (where $-e$ is the charge on the electron) and $-(1/3)e$ to the other two, the charges on all the known mesons and baryons came out correctly. But the idea of fractional charges was fairly repulsive to physicists of the day; in his original paper, Gell-Mann even wrote that "a search for stable quarks of charge $-1/3$ or $+2/3$ at the highest energy accelerators would help to reassure us of the nonexistence of real quarks" (*1*, p. 215). After several years of fruitless searches (*3*), most particle physicists agreed that although quarks might be useful mathematical constructs, they had no innate physical reality as objects of experience.

The First MIT-SLAC Experiments

The first electron-proton scattering experiment at SLAC, in which electrons with energies up to 20 GeV (1 GeV equals 1 billion electron volts) recoiled elastically from the proton (that is, without breaking it up), gave no evidence for quark substructure (*8*). The cross section, or probability, for this process continued to plummet—approximately as the 12th power of the invariant momentum transfer from electron to proton—much as had been observed earlier in the decade at lower energies. This behavior was generally interpreted as evidence for a soft proton lacking any core; it was commonly thought that the existence of such a core would have slowed the rate at which the cross section decreased.

In the next experiment, performed in late 1967 by the MIT-SLAC collaboration, electrons rebounded inelastically from protons (*9*); the energy imparted to the proton either kicked it into a higher energy excited state (such as one of the spin-3/2 baryons) or shattered it entirely. In the latter occurrence, known as "deep inelastic scattering," the electron rebounded with much less energy. Theoretical analyses of deep inelastic electron-proton scattering made that year by Bjorken (*10*) suggested that this process might indicate whether there were any constituents inside the proton, but his ideas were not well received initially by the particle physics community.

Inelastic electron scattering was measured with three spectrometers (Fig. 1) in SLAC End Station A that were built largely under Taylor's direction. A beam of electrons with energy E passed through a liquid hydrogen (and later also a deuterium) target (Fig. 2). Electrons that rebounded at a

The author is with the Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309.

preselected angle θ into the spectrometer were momentum-analyzed; those with a scattered energy that fell into a range of about ±2% around a central value E' were directed onto a group of particle detectors that distinguished electrons from a background consisting mostly of pions. For each given set of values E and θ, measurements were made at a series of scattered energies ranging from elastic electron-proton scattering at the highest E' down to deep inelastic scattering at a few giga-electron-volts.

In the first inelastic experiment, which took place in the autumn of 1967, the 20-GeV spectrometer was used to measure electrons that rebounded from protons at an angle of 6°. The raw counting rates were much higher than had been expected in the deep inelastic region, where the electron imparts most of its energy to the proton, but there was considerable disagreement among the MIT and SLAC physicists as to the proper interpretation of this effect. Electrons can radiate photons profusely as they recoil from a nucleus or pass through matter (in this case, the surrounding hydrogen and target walls); such an effect, which can lower their energies substantially, has to be removed from the raw data before one can assess the underlying physics. These "radiative corrections" were very time-consuming and full of uncertainties; they involved measuring cross sections over a large range of E and E' for a each value of θ. After the experimental run was over, a computer program (*11*) was used to deconvolute these data and obtain corrected cross sections at the same kinematics as measured.

When the radiative corrections were completed in the spring of 1968, it became clear that the high counting rates in the deep inelastic region were not due to radiative effects. A plot of the cross section σ versus the invariant momentum transfer to the proton, $q^2 = 2EE'(1 - \cos\theta)$, showed that the probability of deep inelastic scattering decreased much more slowly with q^2 (also written Q^2) than that for elastic scattering (Fig. 3). A way to interpret this unexpected behavior was that the electrons were hitting some kind of hard core inside the target protons. In hindsight, such an observation paralleled the discovery of the atomic nucleus by Ernest Rutherford (*12*), in which the probability of large-angle α-particle scattering from gold atoms was found to be far larger than had been anticipated based on J. J. Thomson's "plum pudding" model of the atom. At the time, however, there were a few other possible interpretations of the inelastic electron-scattering data (*13*) that had to be excluded before one could conclude that the MIT-SLAC group had indeed found evidence for constituents inside the proton.

Scaling and the Parton Model

In April 1968, at the suggestion of Bjorken, Kendall plotted the quantity νW_2 versus the variable ν/q^2, where $\nu = E - E'$ is the energy lost by electrons in the act of scattering and W_2 is known as a "structure function" of the proton. In the first Born approximation, wherein a single virtual photon mediates the electromagnetic interaction between the electron and proton (Fig. 4), there are two such structure functions, W_1 and W_2; they contain all of the information that can be obtained about the proton from unpolarized electron scattering

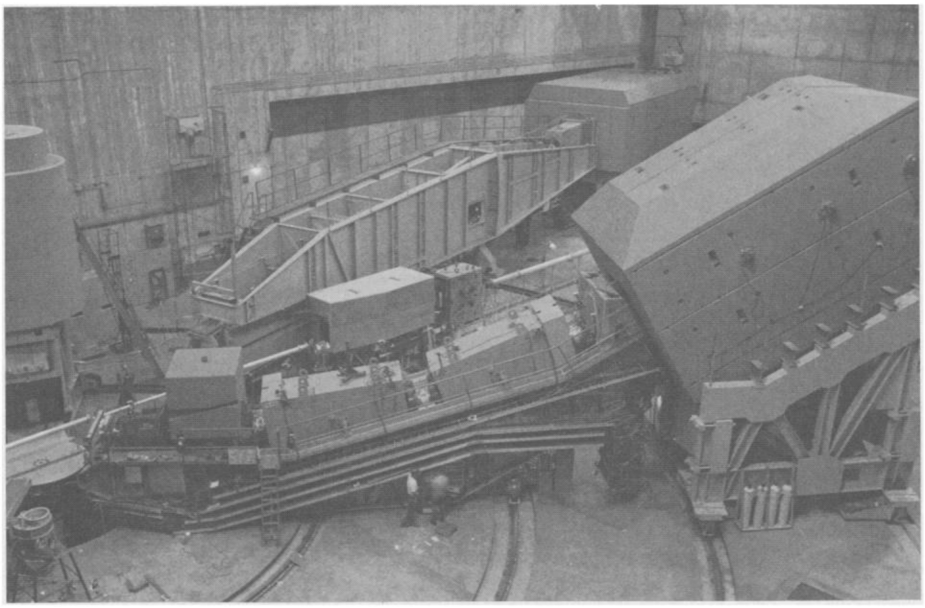

Fig. 1. The End Station A spectrometers used in the MIT-SLAC experiments. A beam of multi-GeV electrons passed through targets on the pivot at the extreme left of this photograph; scattered electrons were momentum-analyzed and discriminated from other particles by the 1.6-GeV (far left), the 8-GeV (foreground), or the 20-GeV (rear) spectrometers. [Photo courtesy of SLAC]

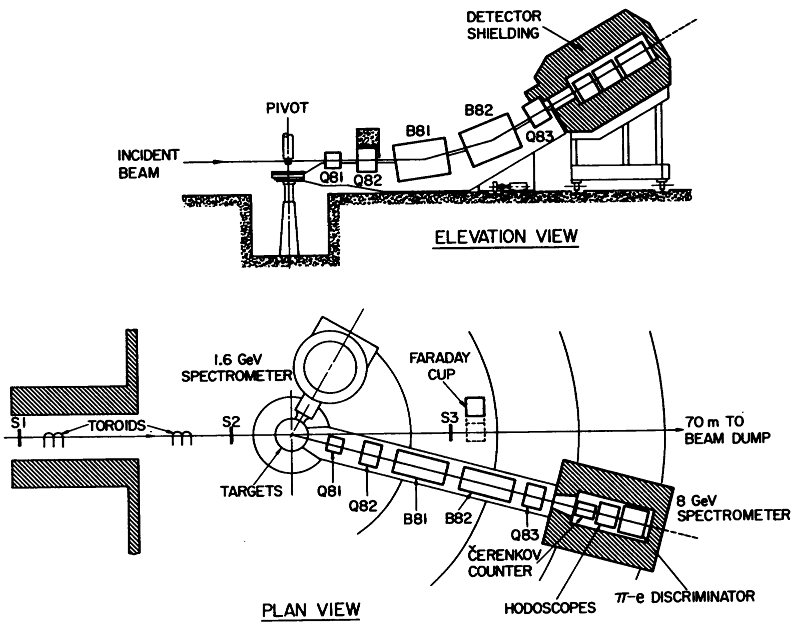

Fig. 2. Schematic diagram of the experimental setup for the inelastic electron-scattering experiments performed with the 8-GeV spectrometer (*27*). Five electromagnets—two dipoles (B) and three quadrupoles (Q)—bent and focused particles scattered by the targets onto a series of detectors inside a shielded cave. The intensity of the electron beam was measured by two toroidal charge monitors, which were periodically calibrated against a Faraday cup.

and are related to the cross section σ by

$$\sigma(E,E',\theta) = \frac{4e^4E'^2}{q^4}\left[W_2(\nu,q^2)\cos^2\frac{\theta}{2} + 2W_1(\nu,q^2)\sin^2\frac{\theta}{2}\right] \quad (1)$$

Clearly, W_2 dominates the cross section at small angles, while W_1 determines large-angle scattering. Making two extreme assumptions about the ratio W_2/W_1, Kendall extracted values of W_2 from the 6° cross-section data, obtaining a graph like Fig. 5. As Bjorken predicted (*14*), the data appeared to "scale"—that is, they fell along a single curve $F_2 = \nu W_2$ that is a function of only the ratio ν/q^2, and not ν and q^2 independently, despite the fact that the cross sections had been measured at several different energies.

The physical significance of this curve and the scaling behavior became clearer in August 1968 when Feynman interpreted them in terms of a model in which protons were composed of generic pointlike constituents he called "partons." In this model (*15*), scaling arose naturally because high-energy electrons rebounded elastically from charged, pointlike partons; he recognized that the universal function F_2 was the momentum distribution of the partons, weighted by the squares of their charges, when plotted versus a variable $x = q^2/2M\nu$, where M is the mass of the proton. Note that x is actually the inverse of Bjorken's variable; it represents the fraction of the proton momentum carried by the struck parton when viewed in what Feynman called the "infinite momentum frame"—essentially the reference frame in which the electron is at rest and the proton is speeding toward it.

In his model, Feynman did not advocate any specific quantum numbers for the partons; they could have whatever charges, spins, and other properties were consistent with the MIT-SLAC data. On the basis of these ideas, other physicists soon formulated more specific parton models in which the partons were interpreted as quarks (*16*) or as bare, pointlike nucleons and mesons (*17*). The spin of the partons could be determined (*18*) from the behavior of the quantity $R = \sigma_L/\sigma_T$, the ratio of the proton's tendencies to absorb virtual photons that are polarized longitudinally (that is, along their direction of motion) or transversely; R is related to W_1 and W_2 according to

$$R = \frac{W_2}{W_1}\left(1 + \frac{\nu^2}{q^2}\right) - 1 \quad (2)$$

Competing theories (*19, 20*) could also account for the observed scaling behavior without invoking proton constituents; thus, further, more detailed measurements of deep inelastic scattering were necessary before any firm conclusions could be drawn about what was happening inside the proton.

Indeed, such measurements were already well under way at SLAC by the end of the year. In August, the MIT-SLAC physicists obtained cross sections at 10° with the 20-GeV spectrometer, and that autumn they used the 8-GeV spectrometer for measurements at 18°, 26°, and 34°. In addition to determining inelastic electron-proton cross sections over a much wider kinematic range, these experiments allowed the group to extract both structure functions, and hence R, at selected kinematic points where data were available for several angles. After radiative corrections had been applied, preliminary data for these angles were presented at the Liverpool Conference (*21*) in the summer of 1969. It became obvious there that the measured values of R were small, consistent with the charged partons being spin-1/2 particles and completely at odds with models based on vector meson dominance (*19*), which required R to be large and proportional to q^2.

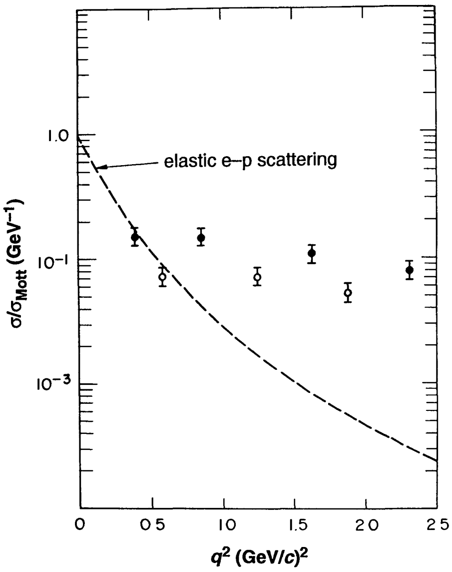

Fig. 3. Cross sections for inelastic electron-proton scattering measured at 6° in the first MIT-SLAC experiment, normalized with those expected for Mott scattering from a point proton (*4*). The data points are given for two values of *W*, the invariant mass of the unobserved final-state hadrons [(●), 2 GeV; (○), 3 GeV].

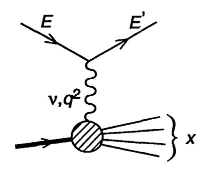

Fig. 4. Feynman diagram for inelastic electron scattering from a proton or neutron in the first Born approximation, in which only a single virtual photon, with energy ν and invariant momentum transfer q^2, is exchanged.

The data from the 6° and 10° inelastic electron-proton scattering experiments were published in two papers (*22*) that rank among the most highly cited in particle physics for 1969; the 18°, 26°, and 34° data were published a few years later (*23*) but were widely available well before that. Graphs of νW_2 and $2MW_1$ versus $\omega = 1/x$ (Fig. 6) showed that both structure functions scaled, within the accuracy of the data, consistent with expectations based on parton models. Although vector dominance had essentially been ruled out, theoretical models based on Regge exchange (*20*) could still account for the general features of the data. And while partons seemed in good shape, little could be said about their physical properties, other than that the data for $R = \sigma_L/\sigma_T$ favored a value of spin-1/2 for charged partons.

Further MIT-SLAC Experiments

More detailed studies of the nucleon's interior came during the next round of MIT-SLAC experiments, in which inelastic electron scattering from both protons and neu-

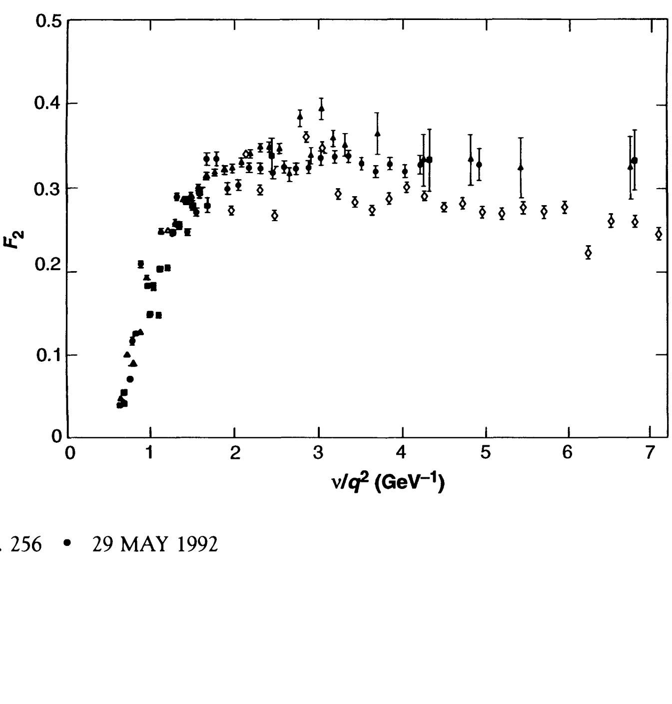

Fig. 5. Values of the proton structure function $F_2 = \nu W_2$ derived from the inelastic cross section measured in the first MIT-SLAC experiment (*22*) [(■), 6°, 16 GeV; (▲), 6°, 13.5 GeV; (●), 6°, 10 GeV; and (◇), 6°, 7 GeV]. As anticipated by Bjorken, the data appeared to be a universal function of the ratio ν/q^2—especially when the low-q^2 data at E = 7 GeV were excluded from the sample.

trons was measured—and with substantially greater accuracy. As the probability of electron-parton scattering is proportional to the square of the parton's charge, such a comparison of proton and neutron cross sections was designed to help differentiate between the various parton models (*16, 17, 24*) that were being advocated at the time. In the simplest quark-parton model, for example, one in which the proton (*uud*) and neutron (*udd*) contain only three charged quarks, all with the same distribution in momentum, the ratio of neutron to proton cross sections σ_n/σ_p should be 2/3, which is just the ratio of the sums of the squares of the quark charges. In more complicated parton models, this ratio can be different or vary as a function of *x*. In models where the electron scatters diffractively from the nucleon as a whole, σ_n/σ_p was expected to be unity. Measurement of this ratio therefore became one of the principal goals of the second generation of MIT-SLAC experiments, which occurred during the period 1970–73.

Because free neutrons do not exist naturally (they decay within minutes), high-energy electron beams were passed through targets of liquid deuterium, which has a nucleus composed of a proton and a neutron. Measurements made at the same *E*, *E'*, and θ with liquid hydrogen targets allowed subtraction of the proton contribution and extraction of cross sections for electron-neutron scattering. Corrections were made (*25*) for the internal motion of the proton and neutron within a deuteron; these "smearing corrections" amount to a few percent at low values of *x*, where $F_2(x)$ varies slowly, but rise to more than 10% for $x > 0.6$, where $F_2(x)$ falls rapidly with increasing *x*.

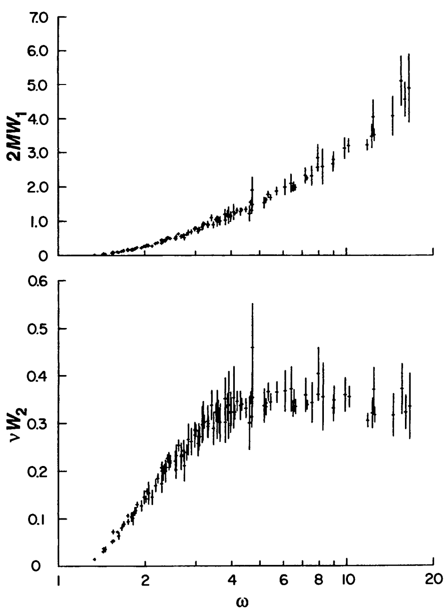

Fig. 6. Values of the structure functions $2MW_1$ and νW_2 ($R = 0.18$; $W > 2.6$ GeV) derived from cross sections measured in the first round of MIT-SLAC deep inelastic electron-proton scattering experiments (*23*). At the level of accuracy attained in these experiments, both structure functions appear to scale in the variable $\omega = 1/x$.

The first experiment with both proton and deuteron targets was done in early 1970, with the 20-GeV spectrometer set to detect electrons scattered at 6° and 10° (*26*); a second experiment later that year used the 8-GeV spectrometer at angles of 18°, 26°, and 34° (*27*). Further measurements were made in 1971 with the 8-GeV spectrometer at 15°, 19°, 26°, and 34° (*28*) to improve the accuracy of the data at $x > 0.5$. These experiments revealed that the ratio σ_n/σ_p itself scales and that it is close to 1 at *x* near 0 but falls to about 0.3 at the highest values of *x* for which it can be reliably extracted (Fig. 7). The data excluded purely diffractive models, which cannot account for a ratio less than unity.

Within the quark-parton picture, the ratio has to fall between 0.25 and 4.0—depending on the momentum distribution of the *u* and *d* quarks within the proton and neutron (*29*). Although the MIT-SLAC data come close to the lower limit of this range as *x* approaches 1, such a behavior is possible if the odd quark (the *d* quark in the proton and the *u* quark in the neutron) is the only charged parton that is ever found carrying almost all of the nucleon's momentum. The fact that σ_n/σ_p approaches 1 when *x* is near 0 can also be explained within quark-parton models (*16, 24*); at low values of *x*, the dominant process is electron scattering from a "sea" of low-momentum quark-antiquark pairs that is the same in both the proton and neutron. At high *x*, however, an electron usually encounters the "valence" quarks, which differ for the two cases.

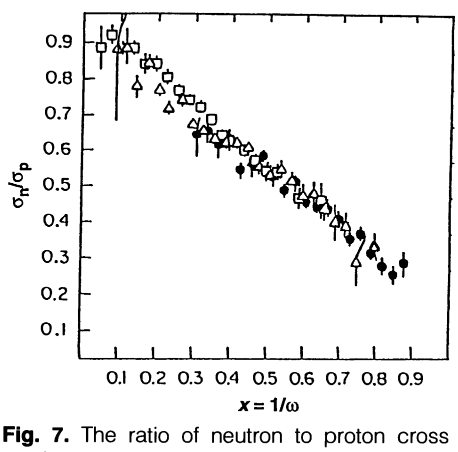

Fig. 7. The ratio of neutron to proton cross sections, as measured in three separate MIT-SLAC experiments (*28*) [(●), 15°, 19°, 26°, and 34°; (△), 18°, 26°, and 34°; and (□), 6° and 10°]. These data appeared to be a single function of $x = 1/\omega$ that decreased from unity at $x = 0$ to about 0.3 at the highest values of *x* measured.

Integrals over the structure functions called "sum rules," which could then be evaluated with the improved data sets, gave added confidence in the quark-parton model (*4*). Because the structure functions represent sums over the various probabilities of an electron encountering each kind of parton (multiplied by the square of its charge), specific parton models give definite predictions for these sum rules. Fractional charges were favored by the data, but certain sum rules still came in about half as big as was expected based on a simple three-quark model of the proton. More complex models incorporating neutral "gluons" to mediate the force binding quarks (*24*) were compatible with the data (*4*) if the gluons carried about half of the proton's momentum.

Combined analyses of all of the data from the second-generation experiments (*30, 31*) allowed extraction of $R = \sigma_L/\sigma_T$ and the two structure functions—for the proton, deuteron, and neutron—with substantially greater accuracy than previously possible. The observation that *R* was the same for all three cases allowed physicists to interpret the cross-section ratio σ_n/σ_p as the ratio of structure functions, too. In each case, the magnitude and behavior of *R* were found to be consistent with partons being spin-1/2 particles—as expected if they were quarks. The more detailed investigations of scaling that also became possible with the improved data revealed that the structure functions had little or no variation with q^2 for selected values of $x < 0.3$ but that they decreased slightly with increasing q^2 at higher values of *x* (*31*). Such a slow falloff had been anticipated in parton models that included gluons (*32*); a cloud of gluons surrounding the charged partons was thought to give them a kind of structure that led naturally to small violations of scaling, as observed.

In 1973 the SLAC group made yet another series of inelastic electron-scattering experiments at angles ranging from 10° to 60° with the 20-GeV spectrometer and the 1.6-GeV spectrometer (which until that time had been used only for counting recoil protons). The results of these measurements (*33*) confirmed the violations of scaling found in the earlier analysis and extended it to the higher values of q^2 that could be attained at the larger angles.

Other Experimental Evidence for Quarks

By 1973, experimental and theoretical developments had produced a coherent picture of the nucleon as composed roughly equally of fractionally charged quarks plus neutral gluons. In this picture there are three valence quarks—*uud* in the proton and *udd* in the neutron—that dominate the process of electron scattering at high x, plus a sea of quark-antiquark pairs—$u\bar{u}$, $d\bar{d}$, and $s\bar{s}$—that is essentially the same for the two nucleons and is largely responsible for scattering at low x. Although electrons cannot interact directly with gluons because they have no electric charge, their existence was still necessary to account for features of the data that could not otherwise be explained in a simple quark-parton model.

This picture soon received additional support from another quarter—inelastic neutrino-nucleon scattering experiments (*34, 35*) at the European Center for Particle Physics (CERN) near Geneva. Neutrinos are point particles like the electron and muon, but they have no electric charge and interact only through the weak force. This force, which is also responsible for the phenomenon of nuclear β-decay, is so feeble that most of the time neutrinos speed right through nuclei (and nucleons, too) without ever interacting. But by placing large quantities of matter in the way of intense, high-energy neutrino beams, the CERN physicists began to detect the rare events in which a neutrino did strike a nucleon, causing its recoil or breakup. As was the case in the MIT-SLAC experiments, the probability of inelastic scattering was far larger than had originally been expected on the basis of soft, extended models of the nucleon.

Reanalysis of data taken at CERN in the 1960s with a bubble chamber filled with liquid freon and propane revealed that the cross section for neutrino-nucleon scattering was proportional to neutrino energy (*34*), as expected if nucleons contained pointlike constituents (*14*); other characteristics of the data favored a value of spin-1/2 for these constituents. More detailed measurements (*35*) made in the early 1970s with the heavy-liquid bubble chamber Gargamelle allowed CERN physicists to extract structure functions from the cross sections for neutrino scattering. These functions coincided (albeit with much larger errors) with the data for $F_2(x)$ that had already been measured in the MIT-SLAC experiment multiplied by 18/5, a factor specified by the quark-parton model (see Fig. 8). Such good agreement provided strong evidence that the partons being hit by electrons and neutrinos carried the fractional electric charges expected of quarks. In addition, sum rules evaluated with these neutrino structure functions (*36*) showed that there were three valence quarks and that only about half of the nucleon's momentum was carried by charged partons, leaving the other half to be carried by neutral gluons.

During the early 1970s, experiments involving colliding-beam techniques, in which two beams of subatomic particles circulating in storage rings repeatedly clash at a few crossover points, also began to provide evidence for the existence of partons. At the ADONE machine in Frascati, Italy, electrons collided with their antiparticles (called "positrons") at combined energies as high as 3 GeV. The probability for such electron-positron collisions to yield hadrons was found to be far larger than had been expected on the basis of models in which hadrons were soft, extended objects (*37*). But these results could be readily accommodated in the quark-parton model as long as the quarks—and the gluons, too—carried an additional property called "color," which was needed anyway for several theoretical reasons (*38*).

Further evidence for the existence of partons within the nucleon came from proton-proton collisions at the very high center-of-mass energies that became possible with the start-up of the CERN Intersecting Storage Rings in 1971–72. Some of the first experiments on this radical new machine (*39*) discovered that far more particles were produced at large angles than could ever have been accommodated from the Regge exchange processes that were then thought to dominate purely hadronic scattering. Counting rates at wide angles were orders of magnitude larger than expected. But this anomaly had a ready explanation within Feynman's parton model (*15*), whether or not the partons were taken to be quarks: the wide-angle scattering was the natural result of close encounters between two partons, one in each proton.

By 1973, then, there was substantial evidence for nucleon constituents in four different kinds of high-energy scattering experiments—electron-nucleon, neutrino-nucleon, electron-positron, and proton-proton—in which three different kinds of forces were involved: the electromagnetic, the weak, and the strong forces. There was also fairly solid evidence that these constituents had the quantum numbers expected of quarks. But quarks had never been convincingly observed in nature, despite continuing efforts to find them. This was the principal quandary remaining about the quark-parton idea. If quarks really existed inside nucleons, why had nobody ever observed any come out?

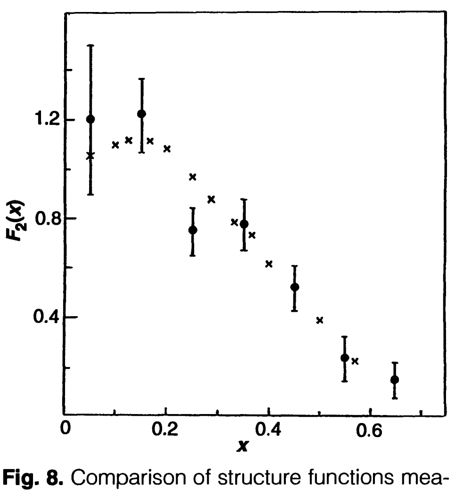

Fig. 8. Comparison of structure functions measured in deep inelastic neutrino-nucleon scattering experiments on the Gargamelle heavy-liquid bubble chamber with the MIT-SLAC data [(●), Gargamelle, $F_2^{\nu N}$; (×), MIT-SLAC, (18/5)F_2^{eN}]. When multiplied by 18/5, a number specified by the quark-parton model, the electron scattering data coincide with the neutrino data.

The Reality of Quarks

The resolution of this quandary came swiftly during the mid-1970s, from both theoretical and experimental quarters (*40*). The unification of the electromagnetic and weak forces within the framework of gauge field theories (*41*) led to their application to the strong force, too. In the summer of 1973 physicists at Harvard and Princeton demonstrated that in certain gauge theories the force between the quarks could become relatively feeble at short distances (*42*), a behavior known as "asymptotic freedom," which could explain why high-energy electrons and neutrinos appeared to be hitting loosely bound quarks inside nucleons. There were expectations that this force also became extremely strong at large distances (that is, comparable to the radius of a nucleon), effectively trapping the quarks—although such long-range behavior could not yet be rigorously derived from the theory.

Experimental support for this theory of the interquark force, which was dubbed "quantum chromodynamics," or QCD, came in gradually during the rest of the decade. One of its key predictions was the occurrence of logarithmic (in q^2) scaling violations, which arose because of the radiation of the gluons needed to mediate this force. Early indications of such behavior were observed in the second-generation MIT-SLAC data (*30, 31, 33*), but there were ambiguities about its appropriate interpretation because of the limited kinematic range of the data. After measurements made at much higher q^2 with beams of muons at the Fermi National Accelerator Laboratory (Fermilab) near Chicago and at CERN (*43*) also revealed scaling violations consistent with logarithmic behavior, the particle physics community grew more confident that QCD was indeed correct. The establishment of a gauge theory in agreement with experiment that could account for the short-distance behavior of quarks and keep them trapped in hadrons helped convince many physicists that quarks actually existed as real, physical particles.

Complete conversion of the physics community came in the mid-1970s, in the aftermath of a remarkable chain of discoveries dubbed the November Revolution. There was no way to explain the J and ψ particles discovered in 1974 at Brookhaven National Laboratory (*44*) on Long Island and at SLAC (*45*) without invoking a fourth—or "charmed"—quark in addition to Gell-Mann and Zweig's original three. Particles in this new family were shown to be combinations of a charmed quark with its antiquark. And in 1975, "jets" of hadrons were seen to emerge from high-energy collisions of electrons and positrons (*46*); detailed analysis indicated that these jets were in fact the footprints of individual spin-1/2 particles, as expected for quarks. With visible evidence for their existence in hand, quarks finally won universal acceptance in the physics community, over 10 years after they had first been proposed.

Further experiments in the late 1970s helped round out the new elementary particle table. In 1976 the same physicists that had discovered the ψ particles at SLAC also identified the τ lepton (*47*)—a charged elementary particle that, like the electron and muon, does not feel the effects of the strong force. In 1977 a fifth kind of quark, dubbed "bottom" or "beauty," was discovered at Fermilab (*48*); a sixth quark, called "top" or "truth," with a mass at least 100 times that of the proton, is now being sought. Visible evidence for gluons was discovered in 1979 at the German laboratory DESY, the Deutsches Electronen-Synchroton, as additional jets of hadrons emerging from electron-positron collisions (*49*). Although important discoveries and measurements were made in the following years, the basic picture of hadrons as composed of quarks and antiquarks bound together by gluons was essentially complete by the end of the 1970s.

Summary

The discovery of quarks was a gradual process that took over a decade for the entire sequence of events to unfold. A variety of theoretical insights and experimental results contributed to this drama, but the MIT-SLAC deep inelastic electron-scattering experiments played the pivotal role. The existence of quarks is recognized today as a cornerstone of the standard model, currently the dominant theory of particle physics. In this theory, all matter is composed of elementary quarks and leptons, and the forces between these particles are carried by gauge bosons such as the photon and gluons. The standard model has been able to accommodate all established subatomic phenomena observed so far. Although further experiments at higher energies may lead to major modifications of this theory, it has weathered all challenges for more than a decade.

REFERENCES AND NOTES

1. M. Gell-Mann, *Phys. Lett.* **8**, 214 (1964).
2. G. Zweig, "An SU3 model for strong interaction symmetry and its breaking," *CERN Report No. TH 412* (CERN, Geneva, 1964); in *Symmetries in Elementary Particle Physics*, A. Zichichi, Ed. (Academic Press, New York, 1965), p. 192.
3. For the status of quark searches by the late 1960s, see A. Zichichi, in *Proceedings of the Heidelberg International Conference on Elementary Particles*, H. Filtuth, Ed. (North-Holland, Amsterdam, 1967), p. 266; see also T. Massam, "The quark hunters' progress," *CERN Report No. 68-24* (CERN, Geneva, 1968). A popular account of these quark searches can be found in M. Riordan, *The Hunting of the Quark* (Simon & Schuster, New York, 1987), pp. 108–120.
4. J. I. Friedman and H. W. Kendall, *Annu. Rev. Nucl. Sci.* **22**, 203 (1972). See also the published versions of the Nobel lectures: R. E. Taylor, *Rev. Mod. Phys.* **63**, 573 (1991); H. W. Kendall, *ibid.*, p. 597; J. I. Friedman, *ibid.*, p. 615.
5. R. W. McAllister and R. Hofstader, *Phys. Rev.* **102**, 851 (1956); M. R. Yearian and R. Hofstadter, *ibid.* **110**, 552 (1958).
6. An excellent general summary of the bootstrap model and SU3 symmetry is given in G. F. Chew, M. Gell-Mann, A. H. Rosenfeld, *Sci. Am.* **210**, 74 (February 1964).
7. M. Gell-Mann, "The eightfold way: A theory of strong interaction symmetry," *Caltech Synchrotron Laboratory Report No. CTSL-20* (California Institute of Technology, Pasadena, CA, 1961); Y. Ne'eman, *Nucl. Phys.* **26**, 222 (1961).
8. D. H. Coward *et al.*, *Phys. Rev. Lett.* **20**, 292 (1968); P. N. Kirk *et al.*, *Phys. Rev. D* **8**, 63 (1973).
9. SLAC-MIT-CIT collaboration, "Proposals for initial electron scattering experiments using the SLAC spectrometer facilities," *Stanford Linear Accelerator Center Proposal No. 4* (January 1966).
10. J. D. Bjorken, in *Proceedings of the 1967 International Symposium on Electron and Photon Interactions at High Energies*, Stanford, CA, 5 to 9 September 1967 (Stanford University, Stanford, CA, 1967), p. 109.
11. The MIT and SLAC groups did separate analyses of the data with two different approaches to the radiative corrections. The MIT corrections program was based on calculations of J. I. Friedman (unpublished); for details, see the theses of Breidenbach and Riordan (*22, 27*). The SLAC group used an approach due to L. Mo and Y. S. Tsai [*Rev. Mod. Phys.* **41**, 205 (1969)].
12. E. Rutherford, *Philos. Mag.* **21**, 669 (1911); H. Geiger and E. Marsden, *Proc. R. Soc. London* **82**, 495 (1909).
13. Space limitations do not permit an adequate discussion of these alternative ideas. They can be grouped into a few major categories, including (i) calculations based on vector meson dominance, especially the work of Sakurai (*19*), in which hadrons called vector mesons helped mediate the interaction between electron and proton, and (ii) Regge pole models (*20*), in which an entire series of hadrons were exchanged. Both were part of the dominant 1960s philosophy of particle physics, which viewed hadrons as soft objects lacking internal structure. There were also models derived from current algebra calculated on the light cone (*14*), which are discussed later in the text.
14. J. D. Bjorken, *Phys. Rev.* **179**, 1547 (1969); see also *ibid.* **163**, 1767 (1967).
15. Feynman did not formally publish these calculations. The best resource to consult for the details of his parton model is R. P. Feynman, *Photon-Hadron Interactions* (Benjamin, New York, 1972). A popular account of his formulation of the parton model and its application to the deep inelastic electron-scattering data is M. Riordan, *The Hunting of the Quark* (Simon & Schuster, New York, 1987), pp. 148–155. See also R. P. Feynman, *Phys. Rev. Lett.* **23**, 1415 (1969).
16. J. D. Bjorken and E. A. Paschos, *Phys. Rev.* **185**, 1975 (1969).
17. S. Drell *et al.*, *ibid.* **187**, 2159 (1969).
18. C. G. Callan and D. J. Gross, *Phys. Rev. Lett.* **22**, 156 (1969).
19. J. J. Sakurai, *ibid.*, p. 981.
20. H. D. I. Abarbanel, M. L. Goldberger, S. B. Treiman, *ibid.*, p. 500; H. Harari, *ibid.*, p. 1078.
21. R. E. Taylor, in *Proceedings of the 4th International Symposium on Electron and Photon Interactions at High Energies*, Liverpool, England, 14 to 20 September 1969, D. W. Braben, Ed. (Daresbury Nuclear Physics Laboratory, Daresbury, UK, 1969), p. 251.
22. E. D. Bloom *et al.*, *Phys. Rev. Lett.* **23**, 930 (1969); M. Breidenbach *et al.*, *ibid.*, p. 935; see also M. Breidenbach, thesis, *MIT Laboratory for Nuclear Science Report No. MIT-COO-2098-635* (1970).
23. G. Miller *et al.*, *Phys. Rev. D* **5**, 528 (1972); see also G. Miller, thesis, *Stanford Linear Accelerator Center Report No. SLAC-129* (1975).
24. J. Kuti and V. F. Weisskopf, *Phys. Rev. D* **4**, 3418 (1971).
25. W. B. Atwood and G. B. West, *ibid.* **7**, 773 (1973); see also A. Bodek, *ibid.* **8**, 2331 (1973) and Bodek's thesis (*27*).
26. J. S. Poucher *et al.*, *Phys. Rev. Lett.* **32**, 118 (1974).
27. A. Bodek *et al.*, *ibid.* **30**, 1087 (1973); see also A. Bodek, thesis, *MIT Laboratory for Nuclear Science Report No. COO-3069-116* (1972); E. M. Riordan, thesis, *MIT Laboratory for Nuclear Science Report No. COO-3069-176* (1973).
28. A. Bodek *et al.*, *Phys. Lett. B* **51**, 417 (1974).
29. O. Nachtmann, *Phys. Rev. D* **5**, 686 (1972); see also (*15*) for a more detailed discussion of these limits.
30. E. M. Riordan *et al.*, *Phys. Rev. Lett.* **33**, 561 (1974); *Phys. Lett. B* **52**, 249 (1974); see also Riordan's thesis (*27*).
31. A. Bodek *et al.*, *Phys. Rev. D* **20**, 1471 (1979); see also R. E. Taylor, in *Proceedings of the 1975 International Symposium on Lepton and Photon Interactions at High Energy*, Stanford, CA, 21 to 27 August 1975, W. Kirk, Ed. (Stanford Linear Accelerator Center, Stanford, CA, 1976), p. 679.
32. M. S. Chanowitz and S. D. Drell, *Phys. Rev. Lett.* **30**, 807 (1973).
33. W. B. Atwood *et al.*, *Phys. Lett. B* **64**, 479 (1976); M. Mestayer *et al.*, *Phys. Rev. D* **27**, 285 (1983); see also W. B. Atwood, thesis, *Stanford Linear Accelerator Center Report No. SLAC-185* (1975); M. Mestayer, thesis, *Stanford Linear Accelerator Center Report No. SLAC-214* (1978).
34. I. Budagov *et al.*, *Phys. Lett. B* **30**, 364 (1969).
35. D. H. Perkins, in *Proceedings, XVI International Conference on High-Energy Physics*, University of Chicago, 6 to 13 September 1972, J. D. Jackson and A. Roberts, Eds. (Fermi National Accelerator Laboratory, Batavia, IL, 1972), vol. 4, p. 189.
36. H. Deden *et al.*, *Nucl. Phys. B* **85**, 269 (1975).
37. V. Silvestrini, in *Proceedings, XVI International Conference on High-Energy Physics*, University of Chicago, 6 to 13 September 1972, J. D. Jackson and A. Roberts, Eds. (Fermi National Accelerator Laboratory, Batavia, IL, 1972), vol. 4, p. 10.
38. Although the story of color and QCD has been largely omitted from this account for the sake of brevity, it is an important corollary to the quark discovery. For more details, consult A. Pickering, *Constructing Quarks* (Edinburgh Univ. Press, Edinburgh, UK, 1984), pp. 207–228; M. Riordan, *The Hunting of the Quark* (Simon & Schuster, New York, 1987), pp. 225–244.
39. B. Alper *et al.*, *Phys. Lett. B* **44**, 521 (1973); F. W. Busser *et al.*, *ibid.* **46**, 471 (1973).
40. Only a brief outline of these developments can be included here. For further, more detailed discussions see A. Pickering, *Constructing Quarks* (Edinburgh Univ. Press, Edinburgh, UK, 1984), pp. 159–300; M. Riordan, *The Hunting of the Quark* (Simon & Schuster, New York, 1987), pp. 225–354.
41. S. Weinberg, *Rev. Mod. Phys.* **52**, 515 (1980); A. Salam, *ibid.*, p. 525; S. L. Glashow, *ibid.*, p. 539;

see also G.'t Hooft, *Nucl. Phys. B* **33**, 173 (1971); *ibid.* **35**, 167 (1971).
42. D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973); H. D. Politzer, *ibid.*, p. 1346.
43. For a review of the situation on scaling violations at the end of the 1970s, see W. S. C. Williams, in *Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energies*, Batavia, IL, 23–29 August 1979, T. B. W. Kirk and H. D. I. Abarbanel, Eds. (Fermi National Accelerator Laboratory, Batavia, IL, 1979), p. 384.
44. J. J. Aubert *et al.*, *Phys. Rev. Lett.* **33**, 1404 (1974).
45. J.-E. Augustin *et al.*, *ibid.*, p. 1406; G. S. Abrams *et al.*, *ibid.*, p. 1453.
46. G. Hanson *et al.*, *ibid.* **35**, 1609 (1975).
47. M. L. Perl *et al.*, *ibid.*, p. 1489; M. L. Perl *et al.*, *Phys. Lett. B* **63**, 466 (1976).
48. S. W. Herb *et al.*, *Phys. Rev. Lett.* **39**, 252 (1977).
49. R. Brandelik *et al.*, *Phys. Lett. B* **86**, 243 (1979); D. P. Barber *et al.*, *Phys. Rev. Lett.* **43**, 830 (1979); C. Berger *et al.*, *Phys. Lett. B* **86**, 418 (1979).
50. I thank J. I. Friedman of MIT for reading two drafts of this paper and J. Toll for support and encouragement during 1991, when I worked in Washington, DC, as Staff Scientist for Universities Research Association, Incorporated. Publication of this work was supported by Department of Energy contract DE-AC03-76SF00515.

New Approaches to Nuclear Proliferation Policy

Joseph S. Nye, Jr.

Nuclear proliferation is not one but a complex of problems. One relates to the collapse of the Soviet Union and its effect on the spread of nuclear weapons and knowledge. Second, Iraq's violation of its Non-Proliferation Treaty obligation has exposed certain weaknesses in the traditional regime of multilateral nonproliferation institutions and treaties. Third, Pakistan's achievement of a nuclear weapons capability in the late 1980s brings the postproliferation question to the forefront in South Asia. There is no single solution to this complex set of problems, but the beginning of wisdom is to build upon the successes of the past, add new policy procedures, and, above all, increase the priority given to the issue. Otherwise, we may be faced with the ironic outcome that the widely welcomed end of the Cold War may increase the prospect of nuclear use.

In the aftermath of the Cold War and the Gulf War, the problem of nuclear proliferation has risen to new prominence. The end of the Cold War has reduced the risk of a large-scale nuclear war, but it has also reduced control by the superpowers. Not only has the disintegration of the Soviet Union removed Soviet control over its client states, but it has also raised the question of how many nuclear states will succeed it. The Gulf War showed that Iraq, in violation of its obligations under the Non-Proliferation Treaty (NPT), had a massive program to develop nuclear weapons. The successful Iraqi deception raised questions about the adequacy of national intelligence efforts as well as of the International Atomic Energy Agency (IAEA) inspection system. Now there are questions about North Korea approaching a nuclear weapons capability.

The 1990s will see three major problems in nonproliferation policy. One is the traditional problem of slowing the rate of spread of nuclear weapons to additional countries such as Iraq and North Korea. The second revolves around what to do after proliferation has taken place in regions such as South Asia and the Middle East. The third set of problems relates to the disintegration of the Soviet Union and its effect on the spread of nuclear weapons and knowledge. Each poses separate problems and questions about appropriate policy goals.

The author is Director of the Center for International Affairs, Harvard University, Cambridge, MA 02138.

Policy Objectives

A policy to slow the spread of nuclear weapons is costly in terms of the friction it can create with other countries; witness the ill will created by the suspension of American economic and military aid to Pakistan. It is not surprising that skeptics raise questions about costs. Some political scientists even argue that nonproliferation policy may have the wrong intent (*1*). If nuclear weapons produced prudence between the superpowers during the Cold War, could they not do the same for other pairs of nations, such as Argentina and Brazil, India and Pakistan, and Israel and its Arab neighbors?

There are several reasons to doubt such general replicability. Statistics show a much higher incidence of governmental breakdown through military coups and civil wars in many of the areas where nuclear weapons might spread. In addition, new nuclear weapons states might not be able to build enough survivable weapons to be confident of assured second-strike capability and thus might increase the risk of preemptive attack by frightened neighbors. Few of the new nuclear powers could develop the elaborate system of command and control, the special safety devices, or the satellite verification that reduced the risk of nuclear war between the superpowers. Nuclear stability between the superpowers involved a long learning process (*2*). Opposition to nuclear proliferation is, therefore, not a question of elitism or racism. Some regional situations might see stable nuclear deterrence, but in many the risks of nuclear instability would be high. As more countries develop nuclear weapons, the probability of their use in war increases, as does the probability of their leakage into unauthorized hands or to terrorist groups.

A second sort of skepticism about nonproliferation policy doubts not its value but its feasibility. With time, technology spreads, and nuclear weaponry is a half-century-old technology. As the aphorism goes, "the horse is out of the barn." But such metaphors do a disservice to clear thinking about policy objectives. It matters how many horses are out of the barn and the speed at which they run. If the policy objective is to prevent any spread of technology, then the situation is hopeless. But if the policy objective is to slow the rate of spread so as to manage the destabilizing effects, there has been considerable success. Nearly 40 countries have the technical and economic capabilities to produce nuclear weapons, but fewer than a quarter of this number have done so. This is a sharp contrast to President John F. Kennedy's 1963 prediction of a world in the 1970s with 15 to 25 nuclear weapons states presenting "the greatest possible danger" (*3*).

The United States built the first atomic bomb in 1945, followed by the Soviet Union in 1949, Britain in 1952, France in 1960, and China in 1964. Israel probably developed its covert capability in the late 1960s, and in 1974 India detonated what it called a peaceful nuclear device. Since then the rate of proliferation has slowed, with only two potential cases. Pakistan probably completed a nuclear weapon in the late 1980s, and some observers believe that South Africa developed the capability to build a bomb in the mid-1980s. In 1991, however, South Africa renounced any ambition to become a nuclear weapons state, adhered to the NPT, and agreed to international inspections.

A number of countries have started but given up nuclear weapons programs, in part because of external pressure, but in large part because of the development of a regime of norms and conventions that have reinforced the attitude against the spread of nuclear weapons (*4*). Libya has been trying to develop nuclear weapons since the