Testing General Relativity: 20 Years of Progress

CLIFFORD M. WILL, *Washington University, St. Louis, Missouri*

FOR MONTHS rumors had been circulating about the test of Einstein's theory, but only a few insiders knew what was happening. The astronomical measurements were difficult, the effect sought after was extremely small, and the analysis of the data was tricky. Many had doubted the effect could really be seen with any certainty. But if successful, the test would be a triumph for the observers and a triumph for the theory. Finally the moment came for the official announcement: Before a crowded auditorium, the experiment team reported that their results matched the prediction of general relativity. The news immediately passed to journalists and was published in newspapers and magazines around the world.

The date could have been November 6, 1919. Then, the experiment was the measurement of the deflection of starlight by the eclipsed Sun, announced at the Royal Society of London by Sir Frank Dyson on behalf of the teams of observers. The result confirming the theory caught the attention of a public weary of the Great War, and it helped make Einstein a celebrity. It was the first in a long line of experimental tests of general relativity.

But in fact the date was December 16, 1978, and the experiment was a measure of the change in the orbital period of the "binary pulsar," PSR 1913 + 16. The pulsar was losing orbital energy at just the rate it should if it were emitting gravitational radiation by the amount general relativity predicted. The announcement was made at the Ninth Texas Symposium on Relativistic Astrophysics in Munich, Germany, by Joseph P. Taylor, head of the team of radio astronomers who made the measurements.

Although this announcement did not cause the same stir as the one in 1919, it was no less important. It represented the climax of 20 years of intensive, high-precision testing of general relativity. During this time the theory's predictions were checked in ways and to levels of precision unheard of during the 40 years after the light-deflection measurement. General relativity passed all tests with flying colors, and many alternative theories fell by the wayside. Coincidentally, Taylor's announcement at the Texas Symposium was a fitting event to open 1979, the centenary year of Einstein's birth.

The story of "experimental relativity" is an important one, not only because it bears on the validity of one of the great intellectual achievements of all time, but also because it illustrates an important aspect of all scientific endeavor: the interdependence of theory and experiment.

In the 40 years between 1919 and 1960, experimental progress in testing general relativity was slow and painful, and the observable effects of the theory were thought to be minimal. At the same time, theoretical progress was slow and generally confined to esoteric, formal questions. But between 1960 and 1980 an explosion of activity took place in both experiment and theory, and the interaction between the two fueled major advances on both fronts.

Astronomy has been intimately involved in this development. Testimony to the

Albert Einstein with W. W. Campbell at Mount Wilson Observatory in 1931. Physicist Max Born once called general relativity "the greatest feat of human thinking about nature, the most amazing combination of philosophical penetration, physical intuition, and mathematical skill." Its full relevance to astronomy, however, has only recently been recognized. Caltech photo.

richness of its contributions are some of the terms associated with what has come to be called relativistic astrophysics — pulsars, black holes, Big Bang, gravitational waves, quasars, gravitational lenses — all of which have both theoretical and observational aspects. Similarly, while theorists worked out observable, testable consequences of general relativity and its competitors, experimenters used new technologies to carry out such tests. The period 1960-80 might thus be called the "decades for testing general relativity."

THE FIRST HALF CENTURY

When he derived general relativity, Einstein was not particularly trying to account for experimental results. Instead, he was driven by theoretical criteria of elegance and simplicity. His goal was to produce a theory of gravitation that incorporated both his 1905 special theory of relativity, which dealt with physics in steadily moving reference frames, and the "principle of equivalence," the proposal that physics in an *accelerating* reference frame is in some sense equivalent to physics in a gravitational field. Once his theory was formulated, however, he had to confront it with experiment. This confrontation was based on what came to be known as the "three classical tests."

One of these tests was an immediate success — the explanation of the planet Mercury's anomalous perihelion shift. Astronomers had long known that, after all perturbing effects on Mercury's orbit were accounted for, there remained an unexplained motion of the point at which Mercury is closest to the Sun. This perihelion shift amounted to about 43 arc seconds per century. Despite the best efforts of celestial mechanicians of the late 19th century, no explanation could be found. General relativity accounted for this effect completely, a result that would go unchallenged until 1967.

The next classical test, the deflection of starlight by the Sun, was such a success that it produced the 1919 media event. The measurements of the deflection, which amounts to 1.75 arc seconds for a light ray that just grazes the Sun, were remarkable achievements for their time, but from today's jaundiced viewpoint, they were only of 20 percent accuracy and were plagued by systematic errors. The observations had to be done during a total solar eclipse so that the star field near the Sun could be photographed. This image was compared with photographs of the same field without the Sun. But unknown scale changes between the eclipse and comparison photographic plates, together with the precarious conditions and bad weather of typical eclipse expeditions, contributed to the errors. Nearly a dozen eclipse measurements between 1919 and 1973 resulted in only small improvements in accuracy.

The third classical test was actually the first proposed by Einstein, eight years before he fully formulated the theory of general relativity. This was the gravitational redshift of light. But despite numerous attempts to measure the redshift of spectral lines from the Sun, or from white dwarf stars such as Sirius B and 40 Eridani B, no reliable confirmation of this effect was possible until 1960. Furthermore, by the late 1950's, theorists Leonard I. Schiff and Robert H. Dicke were arguing that the gravitational redshift was not a true test of general relativity, but instead was purely due to the principle of equivalence.

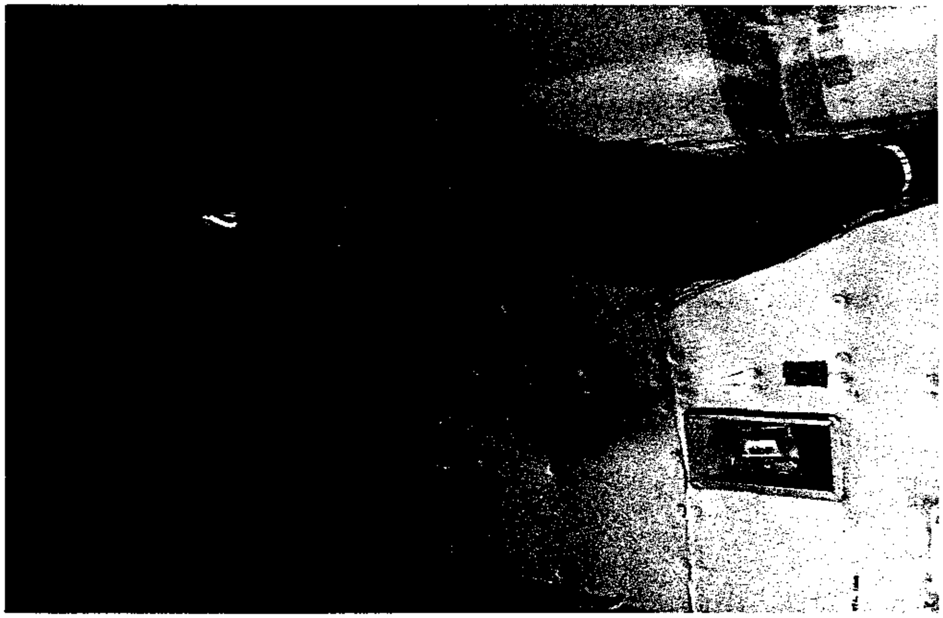
Thus, by 1960, it could be argued that the truth of Einstein's great theory rested upon one test of moderate precision (Mercury's perihelion shift, accurate to about one percent), one test of low precision (the deflection of starlight, no worse than 50 percent), and one inconclusive test that was not a real test anyway (the gravitational redshift).

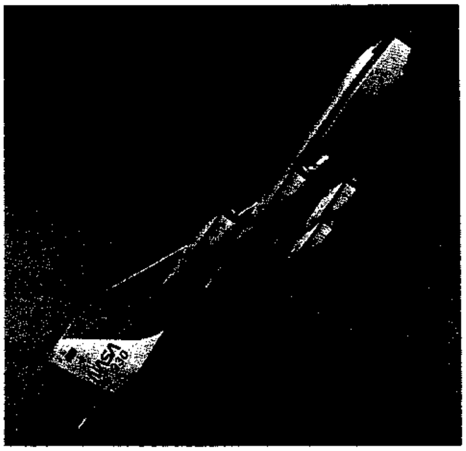
Furthermore, a variety of alternative theories had sprung up over the years and laid claim to equal viability. Because of its weak contact with experiment or observation, general relativity had become a formalistic, sterile subject, cut off from the mainstream of physics.

THE TURNING POINT

In retrospect, the turning point for general relativity occurred during the academic year of 1959-60, though probably no one at the time recognized it as such. During that year, a number of scientific events signaled the start of a new era, one in which the theory would become an important tool for the astrophysicist but would have its validity challenged as never before. It was also to be a time when experimental means would become available to test the theory to its limits.

The first event was the successful recording of a radar echo from Venus, on September 14, 1959. This opened up the solar system as a laboratory for high-precision tests of general relativity. Accurate radar ranging to planets and spacecraft would become an indispensable tool for probing relativistic effects of gravity. Then on March 6, 1960, *Physical Review Letters* received a paper by Robert V. Pound and Glen Rebka, Jr., reporting the first successful laboratory measurement of the gravitational redshift of light. Besides verifying the principle of equivalence, the experiment demonstrated the powerful use of quantum technology, in this case the Mössbauer effect, in gravitational experiments of high precision. That summer the





The principle of equivalence in action. *Left: Sky & Telescope* assistant editor Andrew Chaikin floats weightlessly aboard an airplane following the path of a freely falling object. This is not just an imitation of zero gravity but the real thing; general relativity says that even though the plane is within the Earth's gravitational field, its interior is as truly free from gravity as if it were in deep space. The plane (*right*), which provides experimenters and astronauts-in-training with 25 seconds of weightlessness at a stretch, behaves like Einstein's freely falling elevators — with no disastrous consequences at the end of the ride. Photographs by Otis Imboden (*left*) and NASA (*right*).

Annals of Physics published a paper by Roger Penrose on a "spinor" treatment of general relativity that introduced a new approach to the subject — one using elegant techniques of pure mathematics to streamline calculations in the theory and to help clarify its physical consequences. At the same time, the finishing touches were being put on a new theory of gravity by Dicke and his student Carl H. Brans. The Brans-Dicke theory provided a plausible alternative to general relativity and accentuated the need for high-precision experiments to distinguish between them.

Finally, on September 26, 1960, just over a year after the Venus radar echo, astronomers at Palomar Mountain detected an unusual starlike object at the precise location of the radio source 3C 48. The name "quasar" was soon applied to it and others like it. This and subsequent astronomical discoveries, such as pulsars, binary X-ray sources, and the microwave background radiation, demonstrated important applications of general relativity in astrophysical situations.

After this, the pace of research in the field began to quicken. Numerous advances were made, both theoretical and observational. Among these were measurements of the microwave background, analyses of how the elements could have been formed in the Big Bang, observations of pulsars and black-hole candidates, development of the theory of relativistic bodies and black holes, the theoretical study of gravitational radiation and an experimental program to detect it, the beginnings of a unification of quantum mechanics and gravitation, and, as a nice postscript to the light-deflection measurements, the discovery of gravitational lenses.

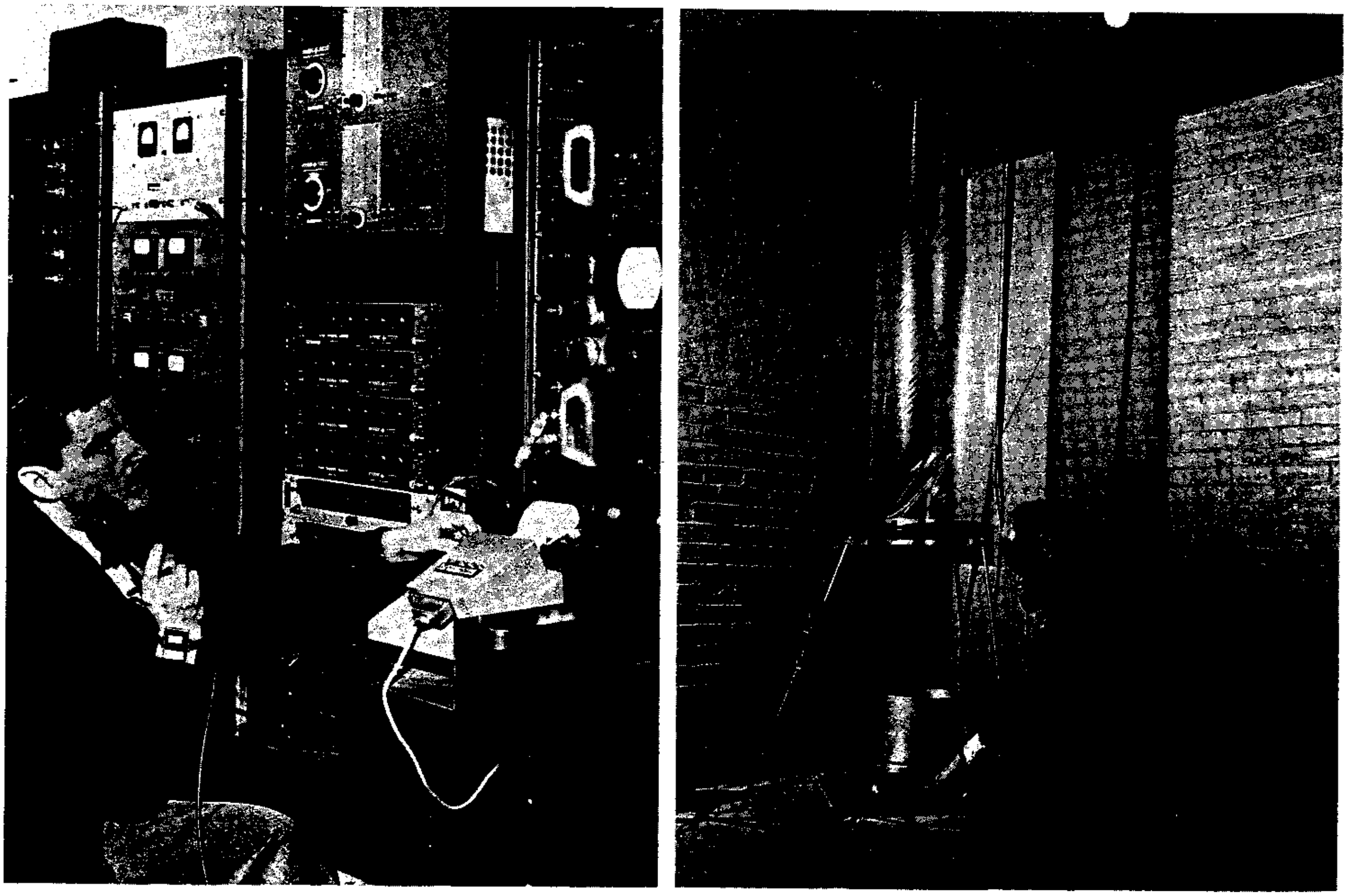
At the same time, systematic, high-precision testing of general relativity became an active and challenging field, with many new theoretical and experimental possibilities. These included new versions of old tests with accuracies unthinkable before 1960, and also brand-new tests of gravitation theory that were discovered theoretically — such as the time delay of light in a gravitational field and the "Nordtvedt effect" in lunar motion. It is to the activity of these two "decades of testing relativity" that we now turn.

TWENTY YEARS OF TESTS

To keep up with the experimental advances, theorists developed a variety of mathematical techniques to analyze the new results and develop suggestions for more experiments. These methods were also used to compare general relativity with competing theories of gravitation, in order to understand the consequences of each.

This approach, pioneered by Dicke and Kenneth Nordtvedt, Jr., in the mid-1960's, allows us to divide the discussion of experiments into four sections, each having a particular theoretical implication. These are (1) tests of Einstein's equivalence principle as the foundation of gravitation theory, (2) the classical tests, (3) tests of a more comprehensive "strong equivalence principle," and (4) tests of gravitational radiation. I will focus on some key tests in each category that best illustrate the progress made possible by both technical and theoretical advances.

Foundations of Gravitation Theory. The principle of equivalence played a central role in mechanics and gravitation from Newton to Einstein. In Newton's view, the principle stated that all objects accelerate at the same rate in a gravitational field, regardless of their mass or composition. Einstein's insight was the recognition that, to an observer inside a freely falling elevator, not only should objects float as if gravity were absent, but also *all* laws of nongravitational physics, such as electromagnetism and quantum mechanics, should behave as if gravity were truly absent.



Robert Pound, left, and Glen Rebka confer during their classic 1960 measurement of the gravitational redshift. Gamma rays from apparatus in the top of a building at Harvard were sent down a tube to a counter in the basement. The tiny extra energy they gained while falling was measured by means of the Mössbauer effect, whereby the nuclei of atoms in a solid emit and absorb extremely sharp gamma-ray spectral lines. The gravitational redshift was still only about one five-hundredth of the line width. Photos courtesy Harvard University.

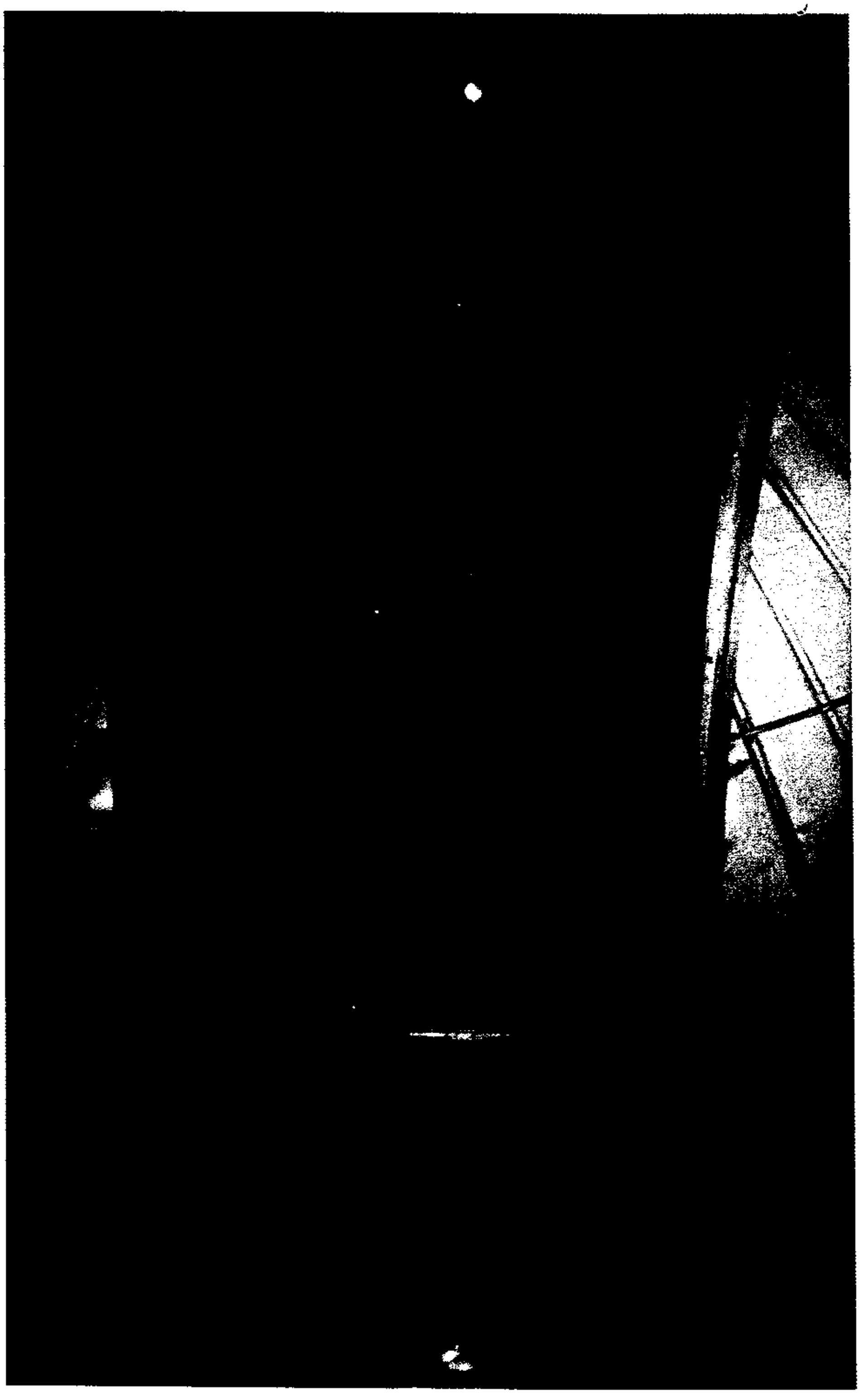
Known as the Einstein equivalence principle, this was a key step, because it implied the converse: that in a reference frame where gravity is felt, such as in a laboratory on the Earth's surface, the effects of gravitation on physical laws can be obtained simply by mathematically transforming the laws from the freely falling frame to the laboratory frame. According to the branch of mathematics known as differential geometry, this is the same as saying that space-time is curved; in other words, that the effects of gravity are *indistinguishable* from the effects of being in curved space-time.

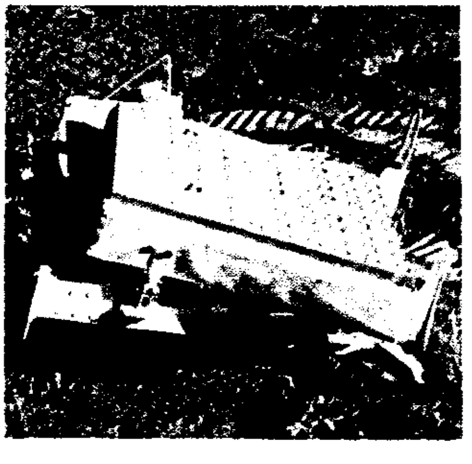
Because this is such a fundamental conclusion about the nature of space and time, it is important to examine its experimental support. A number of experiments can test the Einstein equivalence principle. Here we shall focus on two.

The Eötvös experiment tests the equality of acceleration of different kinds of bodies. Newton himself performed experiments of this sort, but by far the most precise results come from experiments performed at Princeton University in the early 1960's and at Moscow State University in the early 1970's. These experiments adopted the classic method first used by Baron R. von Eötvös around the turn of the century, in which balls of different materials (aluminum and platinum, for instance) are attached to opposite ends of a horizontal bar suspended by a fiber. If the balls accelerate differently toward a distant body, in this case the Sun, the rod will turn in one direction when the Sun is in the east and in the opposite direction when the Sun is in the west. No such rotations were seen to the limits of accuracy of the measurements. This means that different materials fall with the same acceleration to one part in 10^{12}.

As noted above, the gravitational redshift is also a test of the principle of equivalence. Simple arguments based entirely on what would occur in freely falling elevators show that light waves should change frequency when moving up or down through gravity. The first and most famous high-precision test of this was a series of experiments from 1960 to 1965 by Pound, Rebka, and Joseph L. Snider at Harvard University. They found the predicted frequency shift — to a precision of one percent — in gamma rays from radioactive iron when the rays ascended or descended a tower.

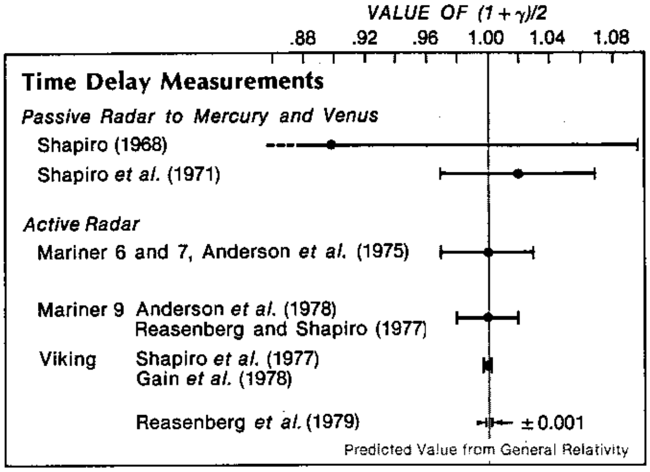
In 1962 and again in 1972, successful (five percent accuracy) measurements of the redshift of the Sun's spectral lines were finally performed. But the best gravitational redshift experiment to date was carried out in June, 1976. An atomic clock was flown on a Scout D rocket to an altitude of 10,000 km, and its frequency was compared by radio signals with that of a similar clock on the ground. After taking into account the effects of the rocket's mo-





Above: A green laser beam is sent to the Moon (partially eclipsed at the time of the photograph) from Lure Observatory, Maui, Hawaii, as part of a lunar ranging experiment. The target is a bank of retroreflector mirrors like the one set up by Apollo 14 astronauts (*right*). Only a few photons at best make the round trip back to the timing instruments at the observatory, but with this technique the Moon's distance can be measured to an accuracy of a few centimeters. Such accuracy is enough to show that the Earth's gravitational field *itself* has mass and exerts a gravity of its own. This confirms a prediction of general relativity and destroys the competing Brans-Dicke theory. Laser-beam photo by Paul Ely.

Many measurements have confirmed that the Sun's gravitational field delays the passage of radio waves. The first two measurements were made by timing radar reflections from planets, the last four by having spacecraft retransmit a signal. Results are expressed as values of the parameter $(1 + \gamma)/2$, which is equal to 1 in the equations of general relativity but not in some of its competitors. The dots are measured values, the bars the observational error.



tion, the observations confirmed the gravitational redshift to 0.02 percent.

Theorists have interpreted these results as convincing evidence that space-time *is* curved, and that the correct theory of gravity must therefore be geometrical or "metric." That does not automatically imply general relativity, however, since it is only one of many possible metric theories. These differ from one another in their predictions of how, and by how much, space-time is curved by matter. To test general relativity itself, other experiments are needed.

Classical Tests. Three such experiments can be called "classical:" the deflection of light by gravity, the time delay of light in a gravitational field, and the perihelion shift of Mercury.

This use of the term "classical" is a break with tradition. I have dropped the gravitational redshift as a classical test, because it is really a test of the principle of equivalence, and have substituted the time delay of light — an effect that is closely related to the deflection of light, since any mechanism that bends light should also "slow it down." In fact, it is a mystery why Einstein or his contemporaries did not think of this effect; it was not noticed until 1964, by Irwin I. Shapiro. Nevertheless, it has given one of the most precise tests of general relativity to date!

The deflection of a light ray that grazes the Sun is different in general relativity than in some competing theories, such as that of Brans and Dicke. The most recent optical measurement of the deflection, during an eclipse expedition to Mauritania in 1973, yielded a result accurate to 10 percent, only a slight improvement over previous attempts. However, the development of very-long-baseline radio interferometry has changed this situation, by providing a means to measure angular separations on the sky as small as 3×10^{-4} arc second.

Coupled with this advance is a series of heavenly coincidences: Each year certain groups of quasars pass close to the Sun, as seen from Earth, including the group 3C 273, 3C 279, and 3C 48. Radio astronomers can measure the relative bending of the signals from the quasars in each group, and have obtained precise values for the deflection. They agree with general relativity to one percent.

Even better accuracy has been obtained for the time-delay effect. A ray that grazes the Sun on a round trip from Earth to Mars, for instance, has its travel time increased by a small amount over what Newtonian theory would predict. In the decade after Shapiro called attention to this effect, several measurements were made using radar-ranging techniques that evolved from the Venus echo work of 1959-60. Two types of targets were employed: planets such as Mercury and Venus, which served as passive reflectors of the radar signals, and spacecraft such as Mariners 6, 7, and 9 and the Viking landers and orbiters, which actively retransmitted the signals. Detailed analyses of the round-trip travel times provided accuracies that improved with each experiment. The latest results, from Viking, yield a value for the non-Newtonian signal delay that agrees with general relativity to one part in a thousand.

Whereas these two tests have undergone enormous improvements since 1960, the third classical test, the perihelion shift of Mercury, has in some sense regressed, and is now in worse shape than ever. To be sure, the measured value of the anomalous shift of about 43 arc seconds per century has improved in accuracy to about half of one percent. But its interpretation has become clouded by the possibility that the Sun is slightly oblate. If the Sun is not exactly spherical, its distorted gravitational field will contribute to the perihelion shift.

Centrifugal flattening due to the Sun's known rotation should have a minuscule effect. But measurements of the Sun's oblateness in 1967 by Dicke and H. Mark Goldenberg led to an inferred value that would contribute three arc seconds to the perihelion shift. This highly publicized result put general relativity in jeopardy (and coincidentally supported the Brans-Dicke theory, which predicted a relativistic contribution of about 40 arc seconds per century). Observations since then have yielded smaller values for the nonspherical component of the Sun's gravitational field, but these are still uncomfortably large. Recently, a very tiny value has been inferred by Henry Hill and collaborators from solar models constructed to fit the observed spectrum of solar oscillations. Because these results involve inferences from solar models, they have been surrounded by controversy and disagreement.

Clearly what is needed are direct measurements of the nonspherical components of the Sun's gravitational field. Such measurements could be provided by Starprobe, a mission currently under study by NASA in which a shielded spacecraft would fly to within four solar radii of the Sun. Orbital data from such a probe could accurately measure effects hundreds of times smaller than those now under dispute. Starprobe may well be one of the most important relativity space missions of the 1980's.

Tests of the Strong Equivalence Principle. One class of experiments testing metric theories of gravity is analogous to testing the Einstein equivalence principle. A "strong" equivalence principle can be formulated, in which not only laboratory-sized bodies, but also planets and stars — bodies with significant gravity of their own — should all fall with the same acceleration in an external gravitational field. In a freely falling elevator large enough to contain, say, a star or planet, the laws of gravity themselves, as well as the laws of nongravitational physics, should behave as if the external world were absent. This is a much stronger principle than Einstein's original version, and to my knowledge it is obeyed *only* in general relativity. Every other known metric theory of gravity violates it at some level. Thus, tests of the strong equivalence principle are crucial to verifying general relativity.

The possibility that gravitationally bound bodies would fall at different rates in different metric theories

MEASURED PARAMETERS OF THE BINARY PULSAR

Parameter	*Measured Value**
Pulse period	0.05902999527*1* ± 2 sec.
Rate of change of pulse period	$8.63 \pm 0.02 \times 10^{-18}$ sec./sec.
Orbital eccentricity	0.617139 ± 3
Orbital period	27906.9816*1* ± 3 sec.
Periastron advance rate	4°.226*1* ± 7 per year
Rate of change of orbital period	$-2.30 \pm 0.22 \times 10^{-12}$ sec./sec.

*Errors are quoted either explicitly or as errors in the italicized last digit.

was pointed out by Nordtvedt in 1968. The difference in acceleration depends upon the self-gravitational binding energy of the bodies and upon a coefficient whose value is zero in general relativity but nonzero in, for instance, the Brans-Dicke theory.

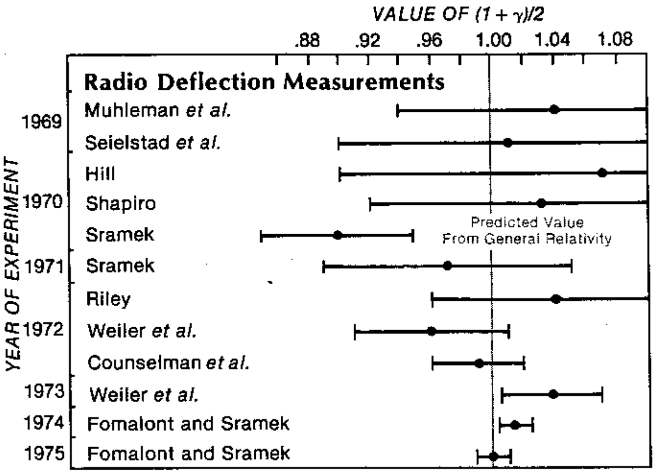
The best test for the existence of this "Nordtvedt effect" is in the Earth-Moon system. The two bodies can be considered as a kind of planetary Eötvös experiment, in which one looks for any difference in the accelerations of the Earth and Moon toward the Sun. If such an effect exists, it would cause an apparent elongation of the lunar orbit oriented along the Earth-Sun axis.

Technical advances in lasers, coupled with space travel, made such an experiment possible. In July, 1969, the first Apollo astronauts on the Moon left behind a small retroreflector — an array of mirrors. One month later the first laser beams were successfully bounced off it from Earth. Since then the Lunar Laser Ranging Program has regularly measured the round-trip travel times of laser pulses sent from several observatories to this and other lunar reflectors, providing accuracies of 30 centimeters in the Earth-Moon distance.

Analysis of six years of data (from 1969 to 1975) yielded *no* evidence of the Nordtvedt effect to an accuracy of a few percent, in agreement with the strong equivalence principle and general relativity. Another way of looking at this result is to say that the Earth and Moon fall with the same acceleration to 7 parts in 10^{12}.

Another important consequence of the strong equivalence principle is that the local gravitational constant G should indeed be a constant, independent of the surrounding environment. There is good evidence (from observations of solid Earth tides and the Earth's rotation rate) that the gravitational constant is independent of the velocity of the Earth relative to the mean rest frame of the universe, and is independent of any particular direction, such as the direction to the massive center of our galaxy.

The evidence is weaker that G is also independent of the mean density of matter in the universe, which decreases as the universe expands (by a few parts in 10^{11} per year). From analyses of lunar occultations, planetary and spacecraft radar ranging, and lunar laser ranging, about all that can be said is that if G varies on a cosmic time-scale, it does so no faster than several parts in 10^{11} per year. While some groups, chiefly involved in lunar occultation studies, claim to have seen statistically significant variations in G at about this rate, other groups claim the data are consistent with constant G within measurement errors. Radar observations of a Mars or Mercury orbiter over a two-year span could reduce the uncertainty to one part in 10^{12} per year.



An advanced version of the classic deflection-of-starlight experiment can be done by measuring the Sun's deflection of radio waves from quasars. Much greater accuracy is possible since positions for these radio sources can be determined far more accurately than for stars. The results of 12 such experiments are expressed here in terms of the same parameter as in the diagram at left. Clearly, relativity wins again. Both diagrams are by the author.

A Test of Gravitational Radiation. The decades for testing general relativity concluded, fittingly, with a confirmation of an important prediction of the theory: the existence of gravitational radiation.

The story begins in the summer of 1974. During a search for new pulsars using the Arecibo radio telescope, Joseph H. Taylor and Russell A. Hulse discovered PSR 1913 + 16. This pulsar proved to be a member of a close binary system with an as yet unseen companion. Its discovery would have been only a mild curiosity (of the first 200 radio pulsars found three are known to be binaries), were it not for two important properties of the system. The two bodies are circling so closely — their orbit is about the size of the Sun, and the period is eight hours — that relativistic effects can be very large. The periastron shift is over 4° per *year.*

Furthermore, the pulsar acts as an extremely stable clock, its pulse period of 59 milliseconds drifting by less than a billionth of a second in four years. By measuring arrival times of radio pulses at Earth, observers were able to determine the motion of the pulsar about its invisible companion and thereby measure many of the orbital elements with amazing accuracy (see the April issue, page 325).

One of the most important predictions of general relativity, and indeed of any reasonable metric theory of gravitation, is the existence of gravitational radiation. Since the late 1960's experimenters around the world have been searching for gravitational waves of astronomical origin. No confirmed detection has been made to date, early reports to the contrary notwithstanding, largely because detectors are still not sensitive enough to detect any but the strongest, and therefore rarest, of the events expected to cause gravitational waves (such as supernovae).

However, the emission of gravitational radiation has another observable consequence, a loss of energy from the system doing the emitting. For the binary pulsar, this energy loss causes the pulsar and its companion to spiral in toward each other and the orbital period to shorten. According to general relativity, the measured orbital elements and masses for the two bodies lead to a predicted change in the orbital period of 2.71 parts in 10^9 (or 75 microseconds) per year. Many alternative theories of gravity predict rates much larger than this.

It was originally thought that measurements of such a small effect would require 10 to 15 years of data, but through crucial efforts to improve electronic techniques at the telescope and to refine the data analysis, Taylor and his team were able to do it in just over four years, in time to open the Einstein centenary with their report at the Texas symposium. Their initial results had 20 percent uncertainty, but subsequent data have improved this to 10 percent. The observed speedup is in complete agreement with the prediction.

During the two decades that closed on the 100th anniversary of Einstein's birth, the empirical foundations supporting general relativity were strengthened as never before. Einstein, of course, was so convinced of the correctness of his theory that he is said to have remarked that he would have felt sorry for the Almighty if the 1919 eclipse measurements had disproved it. Nevertheless, with several patents to his credit, Einstein had a strong feel for experiment. Thus I suspect that he would have been quite pleased and impressed by the "decades for testing relativity."

REFERENCES

For a more exhaustive review of experimental relativity see the author's article "The Confrontation between Gravitation Theory and Experiment" in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, 1979), and at a more technical level, see *Theory and Experiment in Gravitational Physics* by the author (Cambridge University Press, 1981).

Clifford Will has taught physics and performed research at Washington University, Stanford, and Caltech. His research interests include gravitational radiation, black holes, relativistic binary systems, and cosmological perturbations, as well as experimental tests of general relativity.