The Growth of Astrophysical Understanding

A stroll through three millennia of astronomical speculation and discovery reminds us that inspired guesses are not enough. Progress comes primarily from the introduction of new observational and theoretical tools.

Martin Harwit

Perhaps the most remarkable aspect of the growth in our understanding of the universe is that we understand anything at all. Beyond the obvious regularities of the seasons, the Assyrians noted, as early as 700 BC, that the planets appeared to move in a complex semiregular pattern and that solar eclipses were possible only at the new moon, whereas lunar eclipses occurred only at the full moon. But what did all that tell the ancients about the structure of the universe?

Around 250 BC, the Greek natural philosopher Aristarchus of Samos worked out the distance of the Moon and its size. He proposed a method for determining the Sun's distance, but he was able to conclude only that the Sun was much farther away than the Moon and much larger than Earth. That led him to postulate, 18 centuries before Nicolaus Copernicus, that Earth revolves around the Sun.[1]

Aristarchus's theory was largely discredited, especially by Claudius Ptolemaeus of Alexandria. Ptolemy's *Almagest*, which appeared in about 150 AD, dominated Western astronomical thought for a millennium and a half. Ptolemy argued that Earth could not be rotating. Rotation, he thought, would throw anything not firmly attached off the surface, and "animals and other weights would be left hanging in the air." Moreover, Earth's rotation would be so fast that "never would a cloud be seen to move toward the east."[2]

That sounds quaint today, but it wasn't illogical. Ptolemy was a great scientist. The first lesson in astrophysics, however, is that every cosmic phenomenon is governed by competing effects—in this case, gravity, centrifugal forces, and friction. Unless we know the order of magnitude of each, we are likely to draw wrong conclusions.

The observers

When Copernicus revived the notion of a heliocentric system in 1543, he could offer no observational confirmation. The ground for a final resolution had to be prepared by Tycho Brahe (1546–1601), the greatest of the pre-telescope observers. Tycho constructed astronomical instruments more precise than any previously known. Over a 20-year period, he assembled the most accurate, systematic data that had ever been compiled on the positions of the planets.

The young Johannes Kepler, a theorist if ever there was one, dogged Tycho, intent on getting his hands on the data, which the great observer was jealously guarding so he could deduce the orbits of the planets himself. When Tycho was banished in 1597 from his island observatory in Denmark and sought political refuge in Prague, Kepler followed him. But it was not until after Tycho's death that Kepler inherited and began analyzing the data.[3]

One sees parallels to today's theorists impatiently seeking to get an early look at the data from the Wilkinson Microwave Anisotropy Probe's mapping of the cosmic microwave background. The WMAP data were, until just a few months ago, embargoed pending the publication of a full year's set of observations.[4] (See PHYSICS TODAY, April 2003, page 21.) As soon as the data were released, new theoretical analyses began to appear within days on the World Wide Web.

Kepler reduced Tycho's data and arrived at his three laws of planetary motion:

- The planets move in elliptical orbits—rather than in circles and epicycles.
- The rate at which a planet sweeps out area within its orbital ellipse is constant.
- The periods of the planetary orbits increase as the $3/2$ power of their semimajor axes.

The last of these findings was the first quantitative relationship between two observational parameters in astronomy. It constituted what one would call a well-posed question: Why does Kepler's third law hold?

With the advent of the astronomical spyglass in 1609 (the word *telescope* was not coined until the following year), Galileo Galilei quickly discovered an extraordinary new set of phenomena: mountains on the Moon, moons orbiting Jupiter, and the moonlike phases of Venus. To Galileo, those three observations meant that Earth is just one of the planets, all of them orbiting the Sun. For him, that clinched the Copernican theory. The Church, however, forbade Galileo to teach the theory and eventually confined him to house arrest until his death in 1642.

Why did it take until the 17th century for the great discoveries of Kepler and Galileo to come about? Today the answer is clear. Tycho's precision instruments and the spyglass, invented in Holland in 1608 and, a year later, improved by Galileo and pointed at the heavens, provided observational data that had simply been unavailable before.

Although Tycho's instruments gave the best positional data ever assembled, they were still limited by the abilities of the unaided eye, which cannot discern the moons of Jupiter. Galileo's telescopes provided a breakthrough in angular resolution and light-gathering power, a path astronomers are still treading as they build ever larger telescopes and interferometers.

A brief history of instrumentation and its successes illustrates the ability of new instruments to promote astronomical discovery.[5] (See also my article in PHYSICS TODAY November 1981, page 172.) There is a vast range of wavelengths, from the radio domain to the very highest gamma-ray energies, about which Tycho could know nothing. He had only his eyes to rely on, and they merely covered the minuscule visual portion of the electromagnetic spectrum. The naked eye provides a light-gathering aperture of only a few millimeters and a resolution of about an arcminute.

Martin Harwit *is an emeritus professor of astronomy at Cornell University and a former director of the Smithsonian National Air and Space Museum in Washington, DC, where he now lives.*

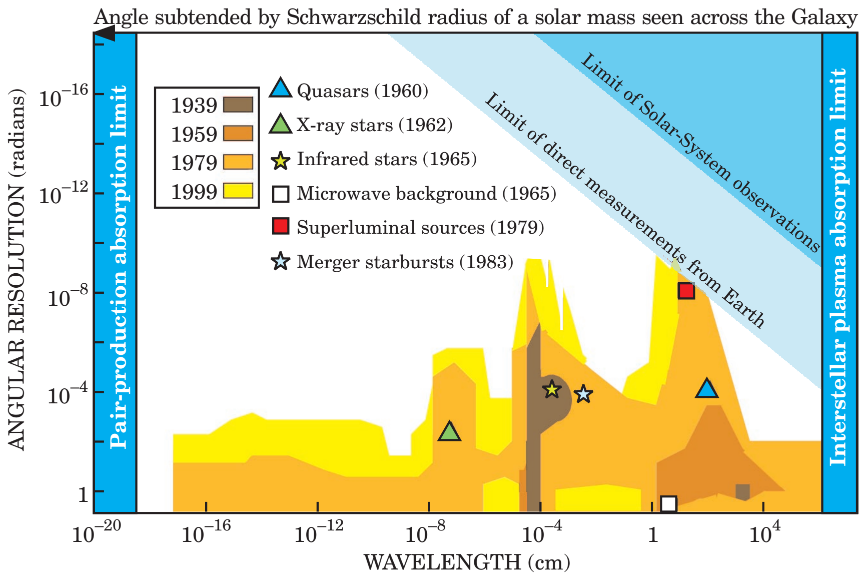


Figure 1. Improving angular resolution of astronomical instruments at 20-year intervals in the 20th century, shown as a function of wavelength. The key discoveries marked on the plot show how promptly new phenomena are revealed when the requisite instrumentation becomes available. Ultimate wavelength boundaries correspond to absorption by interstellar ionized gas and e^+e^- pair production off cosmic-microwave-background photons. Diagonal boundaries mark interferometric baseline limits imposed by the sizes of Earth and the Solar System. An angular resolution of 10^{-18} radians would resolve a 3-km object (the Schwarzschild radius of the Sun) at the far edge of our Galaxy.

Galileo's telescopes increased both light-gathering power and angular resolution by about an order of magnitude. At the start of the 20th century, telescopes had vastly better resolving power and bigger apertures. But they were still restricted to the visible spectrum. Then the development of radar and rockets during World War II opened the radio and ultraviolet regimes to astronomers. To detect UV radiation, rockets had to carry telescopes above the atmosphere. By mid-century, inroads were also being made into the near infrared.

Now, 50 years later, we have access to almost all regions of the electromagnetic spectrum that might be of astrophysical interest. In figure 1, which illustrates progress in angular resolution since the beginning of World War II, we see that resolving power has steadily improved at all wavelengths. But, at some wavelengths more than others, there's still much room for improvement.

The figure also indicates new classes of objects—quasars, x-ray and infrared stars, the cosmic microwave background, superluminal radio sources, and galaxy mergers—revealed by the improving angular resolution at different wavelengths. Most of those discoveries came as huge surprises, which shows that theoretical anticipation had little to do with discovery. What mattered most was the implementation of powerful new observing tools.

Improved angular resolution, however, is not enough. As figure 2 attests, astronomers also needed better timing capabilities to detect such things as slowly expanding supernova remnants, rapid flaring in stars, millisecond pulsars, gamma-ray bursts, quasiperiodic x-ray emission from accretion disks around black holes, and rapid x-ray repeaters.

Then there are phenomena whose discovery demands high spectral resolution, as shown in figure 3. Among these are masers, magnetic stars, and the exquisitely tiny periodic Doppler shifts that reveal stars being tugged by orbiting planets. Polarization measuring capabilities also played an important role.

None of the discoveries highlighted in the figures would have been possible without powerful new instrumentation. Many of those instruments were not originally designed for astronomy; they were mostly hand-me-downs from the military. But it didn't matter. What counted for discovery was instrumental power—adopting the best tools available.

The theorists

What about understanding? What good was it to discover quasars or gamma-ray bursts if you didn't understand the physical processes at work? Genuine astrophysical understanding required a completely different set of tools: theoretical tools.

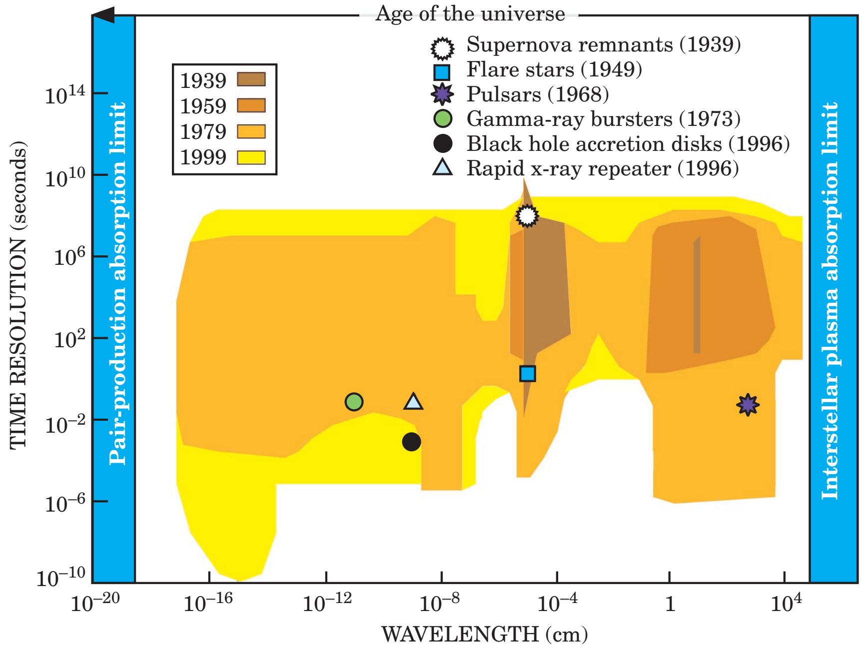
Isaac Newton's discoveries of the laws of motion and gravitation were of a different kind from those of Kepler and Galileo. Newton took into account not only Kepler's laws and Galileo's astronomical observations, but also Galileo's work on projectiles and falling bodies. He conceived that they were all related in some way. To unify celestial and terrestrial phenomena, he had to make use of new theoretical tools. As a young man, Newton had invented the calculus, which now helped him to show that Kepler's laws and the motions of moons could be explained by an inverse-square law of gravitational attraction.

But Newton was not alone in guessing at an inverse-square law. His English contemporary Robert Hooke, best known today for his work on elasticity, came independently to the idea that an inverse-square force law could explain the orbital motion of planets. But Hooke was only able to show that such a law would apply to planets in circular orbits. He lacked the theoretical tools that gave Newton's work its great generality—a universal law of gravitation that held not only for circular orbits, but also for the elliptical orbits of planets and moons, the nearly parabolic orbits of comets, *and* the trajectories of artillery projectiles.[6]

The distinction between having the idea of an inverse-square force and possessing the theoretical tools that quantitatively and convincingly demonstrate it is crucial. I bring this up because the history of science so often alludes to the importance of great ideas. That notion needs to be carefully qualified. In astrophysics, new ideas are afloat all the time. Ideas are, of course, needed. But at critical junctures in the history of astronomy, there is generally an overabundance of ideas on how to move ahead. Supporters of the various ideas debate them vigorously, mostly with no clear-cut outcome. Resolution is usually attained only with the arrival of new theoretical tools that can cut through to new understanding and set a stagnating field in motion again.

Aristarchus and Copernicus are often regarded as the originators of the heliocentric system of the planets. But neither man was able to convince his generation of its va-

Figure 2. Progress of time resolution is important for the discovery of transient and variable phenomena. Format is the same as in figure 1.



lidity. Nor could either make quantitative predictions that would show the idea to be superior to prevailing wisdom.

The convincing proof that our system of planets is indeed heliocentric came from Tycho's precise instruments and painstaking measurements; Galileo's telescope, which gave a far clearer view of planets, moons, and the stellar world beyond; and Newton's dynamics, which made it possible for Edmund Halley (1656–1742) to predict that the comet which now bears his name would be seen again in 1759. Its reappearance, precisely on schedule, created a sensation! This predictive tour de force convincingly showed that not only planets, but also comets, obeyed Newton's universal laws of motion.

Tunneling to the rescue

Let us now jump ahead to the 20th century and look at what it took to arrive at today's conception of the sources of stellar energy. At the end of the 19th century, there was a crisis in attempts to understand what makes the Sun and the stars shine. It was clear that gravitational contraction and chemical energy were insufficient to have kept the Sun shining for a time anything like the apparent age of Earth.

Many ideas were set afloat, ranging from radioactive decay to nuclear reactions of various kinds. But Arthur Eddington had concluded, on hydrostatic grounds, that the temperature at the center of a star would be of the order of 40 million kelvin. That's less than 4 keV, which was thought to be far too low for nuclear reactions to take place.

Then, in 1928, the young physicists Robert Atkinson and Fritz Houtermans became interested in the concept of quantum tunneling that had recently been proposed by George Gamow. Atkinson and Houtermans suggested that protons and electrons might be able to tunnel into a nucleus at a far lower temperature than they would need to leap over its Coulomb barrier.[7] They proposed that a succession of four protons and two electrons could penetrate a helium-4 nucleus to build up unstable beryllium-8, which would decay to two helium nuclei. Thus helium would be the catalyst that lets nuclear energy be generated by the conversion of hydrogen into helium. That idea, having no particular quantitative support, was largely ignored. (See the article about Houtermans by Iosef Khriplovich in PHYSICS TODAY, July 1992, page 29.)

The prolific Gamow had another novel idea. In 1935, he postulated the existence of highly compact, massive neutron cores at the centers of stars. Material falling onto such a core from a much larger surrounding envelope would release enough gravitational energy to keep the star shining for eons.[8]

By 1939, Hans Bethe (shown in figure 4) had the full set of nuclear-physics tools at his command. In his definitive paper, "Energy Production in Stars,"[9] Bethe convincingly showed the importance of the reaction

$$^{1}H + {}^{1}H \rightarrow {}^{2}D + e^{+},$$

with the subsequent addition of two more protons to form ^{4}He. This fusion reaction was, he argued, the primary source of energy in low-mass stars like the Sun. He also laid out the catalytic CNO (carbon-nitrogen-oxygen) cycle for more massive stars with higher central temperatures. Bethe understood the difficulty of accounting for the creation of elements heavier than helium in stars, but he postulated that, somehow, the CNO elements were present in the more massive stars. He knew that the lighter elements—lithium, beryllium, and boron—would all burn in a very short time, and would not be replaced.

In one swoop, Bethe was able to convince the physics

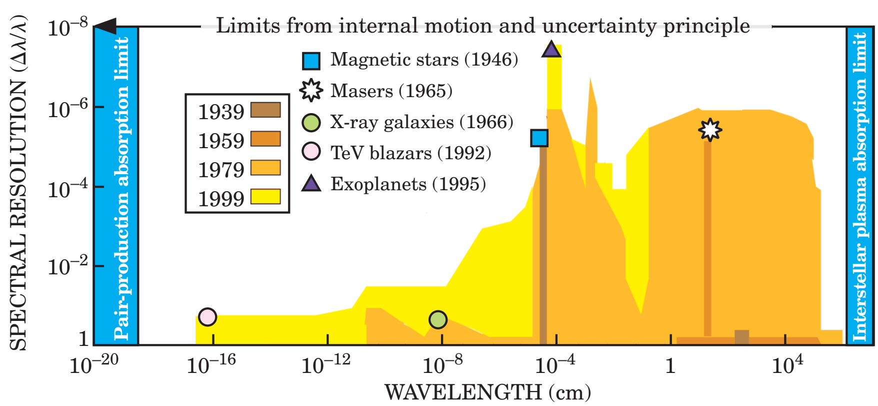


Figure 3. Exquisite spectral resolution, almost at the limits imposed by thermal motion and intrinsic line widths, now makes possible, for example, the discovery of the tiny stellar Doppler shifts due to the periodic tugs of jovian planets. Format is the same as in figures 1 and 2.

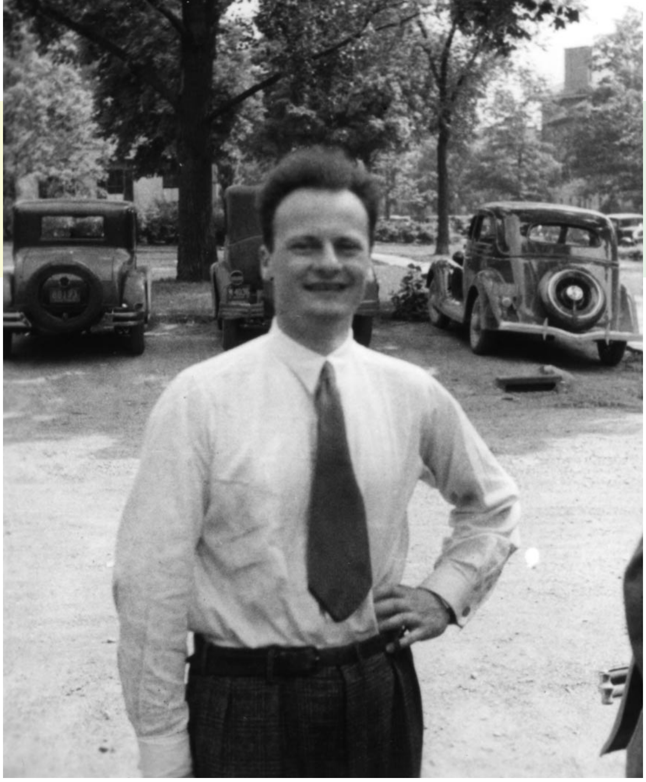


Figure 4. **Hans Bethe** visiting the University of Michigan in 1935. (Courtesy AIP Emilio Segrè Visual Archives, Goudsmit collection.)

and astronomy communities that nuclear reactions fuel the stars. He had the theoretical tools to perform quantitative calculations based largely on data from nuclear-physics laboratories. That made all the difference. Bethe was able to show that the Sun's luminosity is just the energy output one would expect from nuclear reactions at the estimated temperature of the solar core.

Atkinson's general ideas, which he pursued into the mid-1930s, may have been useful. But they were wrong in detail and, by themselves, led nowhere. Gamow's model resurfaced decades later, when Kip Thorne and Anna Żytkow postulated the existence of massive neutron cores at the centers of some evolved giants—stars now known as Thorne–Żytkow objects and still under theoretical investigation.[10]

Too many competing ideas of equivalent merit were floating around in the mid-1930s to settle, by themselves, the stellar-energy question. So, none of those competing ideas made headway. It took the new theoretical tools of nuclear physics, brought to the task by Bethe, to sort out their merits and demerits, and convincingly resolve the problem.

The primacy of tools

We hardly notice any more that our entire understanding of the cosmos rests on Albert Einstein's general theory of relativity. Without the set of theoretical tools he provided, we would be nowhere. Yet Einstein's motivating cosmological idea was to model a universe that was static—neither expanding nor contracting.[11] Fortunately, the tool kit that general relativity provided was far more flexible than was Einstein's world picture at the time. New cosmological models based on general relativity appeared promptly after the discovery of the universal Hubble redshift pattern indicated that the cosmos was indeed expanding.

A further idea of Einstein's, following the discovery of the cosmic expansion, was that the cosmological constant Λ in the equations of general relativity should be excised. Einstein had originally introduced Λ, before the Hubble expansion was discovered, so that the equations would allow a static universe. Fortunately, as we now see it, nobody ever found a proper theoretical justification for discarding Λ. The recent dramatic observation that the Hubble expansion is actually speeding up tells us that Λ may once again be necessary to describe the cosmos. (See the article by Saul Perlmutter in PHYSICS TODAY, April 2003, page 53.) It now appears that about 70% of the mass–energy of the Universe is accounted for by some unexplained "dark energy" that works against gravity on large scales, very much like Λ. An alternative view of this mysterious dark energy that calls it "quintessence," presumes a temporal variability that Einstein's cosmological *constant* would not possess.

Einstein's ideas about a static universe and the cosmological constant were wrong, but the tools he had provided were invaluable. There are many other examples of new theoretical tools in astrophysics that have made a greater long-term impact than mere ideas. One could cite Subrahmanyan Chandrasekhar's introduction of relativistic quantum statistics into the theory of stellar structure in the early 1930s to show the existence of the maximum mass a star can have if it is to escape gravitational collapse (see figure 5 and the article by Kameshwar Wali, in PHYSICS TODAY, October 1982, page 33).

In the late 1940s, Gamow, Ralph Alpher, and Robert Herman (see figure 6) introduced a combination of nuclear

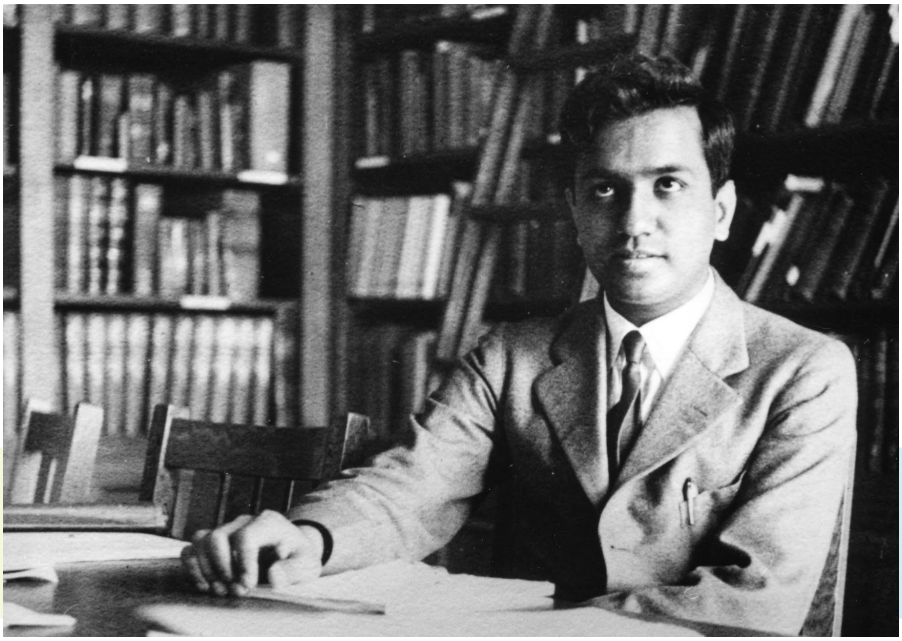


Figure 5. Subrahmanyan Chandrasekhar in the 1930s. (Photo by Dorothy Davis Locanthi, courtesy of AIP Emilio Segrè Visual Archives.)

Figure 6. Robert Herman, George Gamow, and Ralph Alpher (left to right) with their bottle of YLEM, a fanciful primordial form of matter they concocted. Today we would call it quark soup. Alpher and Herman surreptitiously created this montage and sneaked it into a box of slides Gamow had prepared for a talk. When it appeared on the screen, Gamow, after a moment of shock, was very pleased. (Courtesy of Ralph Alpher.)

physics and general relativity into cosmology to determine how hydrogen, deuterium, helium, lithium, and beryllium were created in the first minutes after the Big Bang and to predict the existence of cosmic microwave and neutrino backgrounds.[12] We have not yet been able to observe the predicted neutrino background but, thanks to their work, we have a good idea of the temperature it will exhibit when it's finally detected.

Where are we now?

I started out by saying it was surprising that we understand anything at all. For millennia, the prevailing assumption was that incomprehensible and arbitrary divine forces governed the universe. Today the scientific community subscribes to a very different article of faith: that we will be able to understand the universe and its evolution through the laws of physics. Why do we have this confidence now, when we lacked it for most of recorded history? And is our confidence justified?

To answer the first question, one can say that success has made us heady. Starting with Newton's great insights, we've been making steady progress for three and a half centuries. As we have improved our knowledge of the structure of matter and the nature of energy, we have applied this detailed insight to the quest to understand the cosmos on the grandest scales.

Is progress at this rate likely to continue, or are we about to hit an impenetrable wall? That's hard to say. To gauge how much we have already learned and where we will confront great difficulties, we might consider how much of the universe we have already seen, and how much of that we have understood in terms of the laws of physics.

On the observational side, note the large blank spaces that still remain in figures 1–3. These expanses reflect our inability to observe new phenomena surely lurking out there but requiring better instruments for their discovery. Alternatively, one can compare the number of discoveries attributable to improved instruments with the number independently rediscovered, often by totally unanticipated means, with instruments of quite different kinds. Applying Poisson statistics to this comparison suggests that we have already seen perhaps 30 or 40% of all the major astrophysical phenomena that will ultimately be revealed by photons, cosmic rays, neutrinos, and captured extraterrestrial material.[5]

On the theoretical side, to appreciate how much still remains to be understood, consider that the cosmological parameter Ω_B, the fraction of cosmic mass–energy residing in ordinary (baryonic) matter, is only about 4% (see Michael Turner's column in PHYSICS TODAY, April 2003, page 10). What we have learned about the universe is largely restricted to that 4%. The nature of the other 96%, roughly 73% dark energy and 23% dark matter, is essentially unknown.[4,13] That 4% may be a pretty good measure of our achievements. "Not bad," some might say. "Not all that good," others may respond. We'll just have to see how it all plays out.

What next?

What are the tools that we should expect will yield striking new results over the next few decades? Two new observational regimes, neutrino and gravitational-wave astronomy, are expected to reveal genuinely new phenomena. But there are also less well-known technical areas, largely overlooked in the past, that may yield startling new insights. For one, there is growing awareness that photons can carry not only spin, but also orbital angular momentum. Recently, a group at the University of Glasgow demonstrated an apparatus that can sort out photons carrying different amounts of orbital angular momentum.[14] Another group, at the University of Vienna, has proposed that such photons could provide advantages in optical communication.[15]

Some workers involved in the search for extraterrestrial intelligence are currently looking into optical communication schemes involving visible light. Such schemes might be extended to light that carries orbital angular momentum. An individual photon's orbital angular momentum of $L\hbar$ can encrypt $(1 + \log_2 L)$ bits of information. That could be an energy-saving means for transmitting information across interstellar space. The two possible polarization states of the photon's spin angular momentum, by contrast, can transmit only a single bit of information.[16]

Perhaps the most promising theoretical tool for cosmology at the moment is the development of brane theory (see the article by Nima Arkani-Hamed, Savas Dimopoulos, and Georgi Dvali in PHYSICS TODAY, February 2002, page 35). It postulates an essential difference between gravity and the other fundamental forces: Whereas the strong and the electroweak interactions we can observe are confined on our brane—that is, our four-dimensional spacetime continuum, embedded like a membrane in a higher-dimensional "bulk"—gravity also couples us to neighboring branes. This special pervasive character for gravity is dictated by the general relativistic treatment of gravity as the flip side of geometry. An interesting aspect of this highly speculative theory is that it might explain the low value of the cosmological constant.[17]

Happily, the theory makes predictions that are instigating experimental tests both on small scales and at large accelerators (see PHYSICS TODAY, September 2000, page 22). Tabletop searches for departures from Newtonian gravity at small distances are hoping to measure the distance separating the branes, which could be on the order of millimeters or less. To date, the experiments have found no departure from inverse-square gravity at distances as small as a tenth of a millimeter.[18]

Brane theory comes in many different versions, and they provide tools for doing calculations and making predictions. That's not enough to ensure success, but this bold approach is an exciting step in the struggle to elucidate the 96% of the universe's mass–energy that's still dark. If we're to progress beyond the 4% level of understanding, we have to get beyond the ordinary matter and radiation we already know.

This article is based on a talk given at the American Physical Society's meeting in Philadelphia in April 2003.

References

1. T. Heath, *Aristarchus of Samos, The Ancient Copernicus*, Clarendon Press, Oxford (1913), reprinted by Dover, New York (1981).
2. Ptolemy [Claudius Ptolemaeus], *The Almagest*, R. Taliaferro, trans., vol. 14 in the series *Great Books of the Western World*, Encyclopedia Britannica, Chicago (1955).
3. Kepler's interaction with Tycho and Galileo is well described in A. Koestler's *The Watershed: A Biography of Johannes Kepler*, Anchor Books, Garden City, N.Y. (1960).
4. C. L. Bennett et al. (WMAP collaboration), *Astrophys. J.* suppl. **148**, 1 (2003).
5. M. Harwit, *Cosmic Discovery: The Search, Scope, and Heritage of Astronomy,* MIT Press, Cambridge, Mass. (1984).
6. R. S. Westfall, *Never at Rest: A Biography of Isaac Newton*, Cambridge U. Press, New York (1980).
7. K. Hufbauer, *A Physicist's Solution to the Stellar-Energy Problem 1928–1935*, unpublished manuscript (1985); R. d'E. Atkinson, G. Houtermans, *Z. Phys.* **54**, 656 (1929); R. d'E. Atkinson, *Astrophys. J.* **84**, 73 (1935).
8. G. Gamow, *Ohio J. Sci.* **35**, 406 (1935).
9. H. Bethe, *Phys. Rev.* **55**, 434 (1939).
10. K. S. Thorne, A. N. Żytkow, *Astrophys. J.* **199**, L19 (1975).
11. A. Einstein, in *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, W. Perrett, G. B. Jeffery, trans., Dover, New York (1952), p. 175; Paper originally published in *Sitzungsberichte der Preuss. Akad. d. Wissenschaften* **6**, 142 (1917).
12. R. Alpher, R. Herman, *Genesis of the Big Bang*, Oxford U. Press, New York (2001).
13. D. N. Spergel et al. (WMAP collaboration), *Astrophys. J.* suppl. **148**, 175 (2003).
14. J. Leach et al., *Phys. Rev. Lett.* **88**, 257901 (2002).
15. A. Vaziri, G. Weihs, A. Zeilinger, *Phys. Rev. Lett.* **89**, 240401 (2002).
16. M. Harwit, http://arXiv.org/abs/astro-ph/0307430; *Astrophys. J.* (in press).
17. H. Tye, I. Wasserman, *Phys. Rev. Lett.* **86**, 1682 (2001).
18. J. C. Long et al. *Nature* **421**, 922 (2003). ■