Title
Beyond the bright searchlight of science: The quest for the edge of the world

Author
Trimble, V

Publication Date
1999

Beyond the Bright Searchlight of Science: The Quest for the Edge of the World

Virginia Trimble

Astronomy Department, University of Maryland, College Park, MD 20742
and Physics Department, University of California, Irvine, CA 92697

Abstract. Human efforts to probe the extent of the cosmos clearly date back to pre-literate times, with significant progress occurring in early literate societies, the Renaissance to 18th century, the 19th and 20th centuries, but with some issues remaining for the millennium. The important issues are the size of the earth, the distance to the moon, the earth-sun distance, the distances to the stars, the size of the Milky Way, and the size of the universe as a whole, plus its age. It is perhaps significant that, with one exception, each question was definitively, answered in a later period than the one that first asked it.

INTRODUCTION AND AGES

It has been many generations since those raised in Eurocentric countries have risked anything more than our reputations by asking fundamental questions about the universe. The downside is that we can no longer expect to find definitive answers in the form of seeing the wheels that make the whole cosmic clock work. While estimates of the size of the universe have increased more or less monotonically with time, those of the age have cycled between wide limits. Many preliterate cultures imply in their mythologies a time scale that is 10's to 100's of human generations (and educated individuals can sometimes recite the names of their ancestors back to the supposed beginning).

Of the early literate communities, the writers of Genesis incorporated a total time scale somewhat less than 10,000 years. Byzantium (where this would be the year 7506) thought similarly (with assorted Mesopotamians probably the origin of both chronologies). In contrast, thinkers of the Indian subcontinent arrived at very long times scales, up to 10^{12} years, or perhaps infinite, and this wide range persisted through to the rise of modern science. Applications of basic physics, like conservation of energy and Newtonian gravitation, to astrophysical contexts led 19th century scholars to derive ages ranging from 10^7 yr (the Kelvin-Helmholtz time scale for the sun) up to 10^{12} years (the time needed to produce the observed

CP470, After the Dark Ages: When Galaxies were Young (the Universe at $2 < z < 5$),
edited by Stephen S. Holt and Eric P. Smith

distribution of binary star orbits and of clusters of stars and galaxies, starting with initial conditions that seemed relevant to James Jeans and others). As we approach the end of the 20th century (after yet a different zero point), very few astronomers would argue for an age much outside the range 10-20 Gyr, and this probably counts as progress.

THE SIZE OF THE EARTH

Many mythologies imply an earth that can be circumscribed by humans in years to generations and by gods in a day (consider Ra changing at dawn and dusk from his night boat to his day boat and back again, before sailing either above or below the earth, and Apollo and his chariot). Early maps invariably indicate that the center of the (flat) earth was somewhere quite close to where the map-maker lived.

In contrast, Eratosthenes of Cyrene (-276-195), in the one bit of Greek astronomy that most of us remember, set out to determine the size of a spherical world from the observation that, on 21 June, the sun cast no shadows at noon at Syene (Aswan, roughly), but 7° shadows at Alexandria. Clearly he assumed a distant sun, and we must assume a long-range collaboration, or measurements made over at least two years. With the distance between the two cities given as 5000 stadia, we arrive at a circumference of the earth of about 250,000 stadia. This is generally supposed to be quite accurate, based on the length of some particular representative stadium.

Early Chinese tales tell of a certain Pan-ku who shoved the heavens away from the earth by standing between them and growing 18,000 feet a day for 10 years. The product is about 12,400 miles (not necessarily the same as American statute miles!) while the earth's diameter is about 12,800 km. This is close enough to suggest that the writer had access to the results of some observation analogous to that of Eratosthenes.

FROM EARTH TO MOON

The distance to the moon is the only one of these scales for which the first people reported as having tried to make a measurement came close to the modern answer. The ratio of the size of the earth to the size of the moon's orbit is just the angular diameter of the earth as seen from the moon. Luckily you don't have to go there to look. The angular diameter of the earth's shadow as it crosses the moon during eclipse works just as well, leading to a ratio of about 1/60, which, again allowing for some uncertainty in the length of a stadium is probably pretty close. The idea is credited to Aristarchus of Samos (-310 to -230, since people lived backwards in those days). This is the one case I found where the people who first asked the question got a good answer.

ON TO THE SUN

Aristarchus and his contemporaries reached their level of incompetence in trying to get the solar distance from similar geometric methods. The idea is clever. If the sun is further away than the moon by a factor of 10 or 20 or 30, you will see differences in the lengths of times between the major moon phases (e.g. new to first quarter shorter than first quarter to full). Because the actual ratio is more like 400, the time difference would have been too small for the Greeks to record. But the moon's orbit is actually elliptical, so there is such a phase inequality of a day or so, leading the Greeks to put the sun at about 20 times the distance to the moon. Later, pre-telescopic geometrical attempts did not improve the situation. Thus Tycho and Kepler built their models of the solar system with quite good relative distance but absolute ones much too small.

An opportunity to pin down the absolute scale came in 1672, when Mars made a particularly close approach to the earth, permitting a determination of what is called geocentric parallax. One can do this two ways: with simultaneous measurements of the position of Mars in the sky made from distant points on earth, or from a single point on earth, but waiting for diurnal rotation to carry you over a significant fraction of the diameter. Cassini tried the simultaneous method from Paris with a collaborator in Cayenne, and, later-in the year, the earth rotation method with the assistance of Roemer (who later measured the speed of light from the timing of the eclipses of the moons of Jupiter).

The measurement is a very difficult one, and the close agreement of the three determinations on a parallax for Mars of about 20″ (solar parallax of 10-12″) must have had a contribution from chance and/or bandwagonism. In the first edition of his Principia, Newton actually used an even more erroneous value of solar parallax, 20″ (corresponding to 1 AU = 20,000 earth radii) and so obtained a mass for the sun that was too small by a factor about three.

An accurate value of the geocentric parallax of the sun of 8.4-8.8″ finally came from timing of the transits of Venus in 1769. The modern value is 8.794″, plus several more significant figures, and the main uncertainty in the mass of the sun comes from G, not the distance scale.

NEXT THE STARS

Implicit in many myths and explicit in the writings of Anaximander is a distance to the heavens comparable with the size of the earth. Other Greeks like Aristotle put the stars well outside the realm of the planets. Kepler arrived at a specific number by nesting the known planets in concentric Platonic solids (the ones whose faces are regular polygons) and putting the stars just outside. This leads to a sphere of the stars at about 20 Astronomical Units.

A sun-centered solar system firmly predicts stellar parallax (of the stars relative to your coordinate system, even if they are all at the same distance; otherwise of the stars relative to each other, which is easier to see). Not seeing parallax was one of the very strong arguments in favor of a Ptolemaic, earth-centered model, even with naked eye limits to parallax of a few degrees. Most Greek philosophers balked at the requirement that the stars be more than 100 times the distance to the sun.

By the time of Tycho, carefully-constructed instruments had reduced the upper limit on heliocentric parallax to about 25″, putting the stars at more than 3700 AU. Tycho himself constructed a curious cosmology in which most things orbited the sun, but the sun (and moon) in turn orbited the earth, leading to a prediction of no parallax.

A separate line of thought that implies very large stellar distances is the concept that "stars are suns." Since it is dark at night, despite the presence of thousands of stars in the sky, they must be very much more distant than the sun. Nicolas of Cusa (1401-1464, who became a cardinal and died of natural causes) is generally credited with the first comprehensible enunciation of this idea. He taught in addition that the sun, earth, and stars were all made of the same stuff (in contrast to the Aristotelian quintessence) and pointed out that, in an infinite universe, the center is everywhere and the circumference is nowhere.

The Englishman, Thomas Digges, published similar ideas in his A Perfit Description of the Caelestiall Orbes in 1576, but his drawing shows stars crowded up close together outside the orbit of Saturn. The caption says that (in modern spelling) "this orbe of stars fixed infinitely up extendeth itself in altitude spherically...", and perhaps he would have labeled the drawing "not to scale" if the concept had existed then. The best know propounder of "stars are suns" was, of course, Giordano Bruno, who was unlucky in his choice of time and place in which to propound, but who also seems to have been a rather poor politician and to have been condemned for his attitude toward his fellow churchmen as much as for his attitude toward the universe.

Invention of the telescope led to a new round of attempts to measure parallax. A couple of the unsuccessful ones were nevertheless astronomically useful. William Herschel, believing that all stars had the same intrinsic brightness, focused on close pairs of different apparent brightness, expecting them to be at different distances and so to reveal parallactic motion. What he actually discovered was orbital motion, thereby confirming the earlier claim by John Michell that the large number of close pairs could not be a chance occurrence. He announced the result in 1820. Bradley, in 1829, looked for parallax relative to an earth-centered coordinate system and so discovered the quaintly-named aberration of starlight (the tilt in arrival direction of starlight caused by the earth's orbital motion).

The winners, in 1836, were F. Bessel, Th. Henderson, and F. Struve, looking respectively at 61 Cygni (with a proper motion of 5.2″/yr), Alpha Centauri, and Vega (the latter chosen for brightness, 61 Cyg primarily for its large proper motion, implying a relatively small distance).

Parallax measurements and recognizable binary pairs both demonstrated that real stellar brightnesses cover a wide range, but some feeling that all are about like the sun appears to have carried over into Herschel's star gauging and to have contributed to his small scale for the size of the galaxy. Naked eye stars, or any magnitude limited sample, will be dominated by intrinsically bright ones, and the errors in distance made by assuming that a star is like the sun amount (roughly) to a factor of three for Procyon, 5 for Sirius, 10 for Arcturus, 30 for Canopus, and 100 for Rigel.

THE SIZE AND NATURE OF THE MILKY WAY

William Herschel's star gauging resulted in a somewhat flattened distribution of stars, with total diameter about 6 kpc, the sun very near the center, and one fairly deep gash along the central plane corresponding to the Cygnus rift. The thickness was about one-third the diameter. What all this means is that he was seeing to the edge of the galactic disk on the north-south directions, but not in the plane. And, as a result of neglecting interstellar absorption, he inevitably perceived the edge of the distribution to be at about the same distance in all directions, except toward the galactic center, where obscuration is the largest.

Many other astronomers arrived at rather similar conclusions for the size, shape, and sun-centeredness of the galaxy. Simon Newcomb, in an 1882 book, put regions of nebulae on either side of the Herschel "region of stars or galaxy." Eddington in 1912 put the sun slightly above the galactic plane (it still is), and, most remarkably, Cornelius Easton in 1900 sketched a galactic disk with spiral arms like those of the nebulae, with the center of the spiral galaxy correctly off toward Cygnus, but the sun still solemnly at the center of the coordinate system. Over the same century there was a low-key debate about whether the spiral nebulae might be other galaxies or "island universes." Majority opinion, including that of Lowell and others we still remember, said no; they might, rather, be new solar systems in formation along the lines indicated by the Kant-Laplace hypothesis. The main argument for the spirals being somehow an integral part of the Milky Way was that their positions in the sky avoided the galactic equator.

Harlow Shapley came on the scene near the end of World War I, using RR Lyrae stars in the globular clusters to determine their distances, and tying the scale to galactic plane Cepheids with distance measured from statistical parallax. To a certain extent, his neglect of interstellar absorption and his merging of variable stars of Population I and II into a single set compensated each other, though his "best bet" distance scale within the Milky Way was larger than the current one, putting the sun about 20 kpc (range 13 to 25) from the center of a spherical system with total radius 60 kpc or more. Thus, when Shapley and Heber Curtis faced off in the now well-known debate of 1920, Shapley was advocating a large galaxy, comprising in effect the entire stellar universe, with the sun far from its center. Curtis, in contrast, favored a much smaller galaxy, with the sun near the center, and many other similar stellar systems to be found outside. He noted that many nebulae seemed to have dust lanes in their central planes and suggested that a similar lane in the Milky Way could be the cause of our not seeing spirals in its plane. Jacobus Kapteyn, even as the debate was underway, was carrying out yet another very detailed count of stars as a function of magnitude, color, and position on the sky that would lead once again to a small, sun-centered system, the "Kapteyn universe." Its radius, down to 10% of the central stellar density was about 6 kpc and the half thickness about 2 kpc.

Since 1920, the distance of the sun from the center of the galaxy has undergone several oscillations of relatively small amplitude. Shapley said 20 kpc, Oort 6 and then 10 kpc in about 1926 and 1930. Baade moved us in to about 8.2 kpc in the early 1950s, and the IAU in a resolution (prompted by Oort) back out to 10 kpc in 1965. A later IAU resolution dropped this to 8.5 kpc, and many recent determinations, based on a wide range of stellar types and even the apparent proper motion of Sgr A* as seen from earth have found values of 7-8 kpc. Meanwhile, estimates of the total effective diameter have tended to reach back out to Shapley's 100 kpc and even beyond, with the recognition of outlying globular clusters, the dark halo, and so forth, in our own and other spiral galaxies.

Robert Trümpler is best known for his 1930 recognition of general interstellar absorption (arrived at by comparing distances to open clusters determined from their angular diameters with those determined from star brightnesses). The implications for the Milky Way took a little while to absorb, and, in the same year, he sketched a Milky Way with a spheroid 80 kpc across and centered 18 kpc from us (following Shapley), but, simultaneously, a small, thin Kapteyn universe centered near the sun and with its plane tilted slightly to the plane of the larger system, in the direction we associate with Gould's belt. By 1939, Plaskett could sketch an essentially modern galaxy, with thin plane of stars, still thinner plane of dust, a central bulge or spheroid and globular clusters concentrated toward the center but extending to large radii. He put the sun at Oort's 10 kpc.

The issue of the existence of external galaxies was resolved by Edwin Hubble's 1924-25 recognition of Cepheids in NGC 6822 and M31 soon after. Allan Sandage, however, remembers Milton Humason, toward the end of his life, saying that he, Humason, had marked some tentative Cepheids on a plate of M31 back in about 1921. Humason supposedly showed the plate to Shapley, who moistened a handkerchief and removed the identification markings from the back of the glass. Shapley could have saved himself a good deal of grief a few years later by having forgotten his handkerchief that morning, because, according to Cecilia Payne Gaposchkin, he regarded the Cepheids as "having destroyed his universe."

THE EXTRAGALACTIC DISTANCE SCALE

A number of authors had estimated distances to M31 (the Andromeda Galaxy) before the entire community agreed that this was a meaningful thing to do. For instances, Curtis, Lundmark, and Shapley (who thought the result was ridiculous) all placed it at 200-250 kpc on the basis of bright stars and novae, tied to distances within the Milky Way as then understood. The first attempt to double the distance to M31 came from Opik in 1922 (one of the very few of his important papers to appear in the Astrophysical Journal). He made the assumption that the mass to light ratio of the whole galaxy ought to be the same as that of the solar neighborhood (which had been estimated by Kapteyn and Jeans the same year), and using a published spectrogram of the Nebula taken by Pease on the Mt. Wilson 60″ telescope, came up with a distance of 450 kpc.

The first dozen velocities for "spiral nebulae" came from Vesto Melvin Slipher, working at Lowell Observatory. At least a dozen astronomers before Hubble had looked at them and attempted to derive a "solar motion" or a "solar motion with a K-term" (meaning one that showed constant expansion), or even a "solar motion with a distance-dependent K term", which is what Hubble reported in 1929. Lundmark even allowed for a term in the square of the distance (and, naturally found a better fit, as you nearly always do when adding a parameter). Hubble's work was the first to be taken seriously, in part perhaps because his manner was very convincing, but mostly because he seemed to have much more reliable distances (making use of his discovery of extra-galactic Cepheids) than the other analyzers.

Hubble's distances were consistent, but they were badly wrong. Thus his first paper and subsequent ones through 1936 reported values of what we now call H within 10% of 500 km/sec/Mpc. The period from 1929 to about 1965 was punctuated by large, downward steps in the generally-adopted value of H, each a result of re-interpretations of some critical observation or assumption. As early as 1931, Oort suggested 432 or 260 km/sec/Mpc. A careful recalibration of the Cepheid period-luminosity relation in 1941, by Mineur (who allowed for the interstellar absorption of about 1^m per kpc in V found by Trümpler in his study of star clusters) led to about 320. The German astronomer Behr, in 1951, "pre-discovered" what we now call the Scott effect. He pointed out that assuming a constant luminosity for the brightest galaxy in a cluster would inevitably lead to apparent distances smaller than the real ones by a factor that increases with distance. The point is that, far away, you will recognize only very rich clusters, which contain superluminous galaxies not in your local sample. His work implied H near 240 km/sec/Mpc. Since the effect Behr considered is completely separate from the Cepheid recalibration, a concordance of his work and Mineur's would have taken the Hubble parameter down to about 150 km/sec/Mpc. This did not happen, apparently because nobody brought the two ideas together, and very few people encountered even one of them (Mineur published in French in Annales d'Astrophysique, Behr in German in Astronomische Nachrichten.)

Thus at the Rome General Assembly of the International Astronomical Union, participants were duly astounded by the implications of Walter Baade's not seeing RR Lyrae stars in M31 and by A. David Thackeray's having seen them in the Magellanic Clouds. The distance scale doubled, H decreased, and the universe was suddenly twice its previous age. Humason, Mayall, and Sandage took us down to H = 180 in 1956, and a number of astronomers staked out numbers near 100 between 1958 and 1965. In the next decade, battle lines hardened between a "short" distance scale (H = 100 km/sec/Mpc, G. de Vaucouleurs, S. van den Bergh) and a "long" distance scale (H = 50, A. Sandage, G.A. Tammann). The number of astronomers estimating the Hubble constant has grown even faster than the size of the community as a whole over the past two decades, with 20 or more values published per year. The territory between 50 and about 90 filled in quickly. In recent years, the upper envelope has been moving downward and the lower one holding roughly steady. Thus the median H was 75 in 1992-93 and 60 for 23 papers published in 1998.

THE QUEST FOR LARGER REDSHIFTS

Slipher's first extragalactic spectrogram, in 1912, was that of M31. It revealed a heliocentric velocity of about -300 km/sec. This is made up partly of galactic rotation and partly of the Milky Way - M 31 orbit in the Local Group and is an authentic Doppler shift, not a cosmological redshift. Slipher achieved a "personal best" of +1800 km/sec for NGC 584, but found more distant galaxies beyond the reach of his 24″ telescope (he could, of course, have measured the redshift of 3C273!).

Milton Lassell Humason was the next major collector of redshifts. He quickly doubled Slipher's record, with +3779 km/sec for NGC 7619 and doubled it again to +7800 for NGC 7619 in the Coma cluster in 1929. An improved spectrograph on the 100″ telescope passed the 15,000 km/sec mark by 1934. Humason (unlike Hubble) lived to use the 200″ Hale telescope with some regularity. He retired in 1957, having reached 60,000 km/sec for galaxies in the Hydra cluster several years before. Some of the larger redshifts had been achieved using multi-night exposures, and Humason believed that one could go no further, even with the 200″, given the detectors of the time (photographic plates).

The discovery that strong radio sources were often associated with strong emission lines made a significant difference. Cygnus A, the first optical identification, had a redshift of only 0.06 (measured in 1954 by Baade and Minkowski), but 3C 295 took Minkowski to z = 0.46 shortly before he retired in 1960 (on his very last 200″ observing run, it is usually said).

The first quasar, 3C 273, also came out of the program to identify radio sources, as continued by Maarten Schmidt. The year 1963 saw z = 0.158 for it and z = 0.367 for 3C 48. The record for the largest redshift continued to be held by radio-loud quasars until the mid 1980s. Some steps were steep (to z = 2.01 in one fell swoop to 3C 9 in 1965 – the object that brought Lyman alpha into the visible part of the spectrum and led to tight limits on the amount of neutral hydrogen in intergalactic space). Then came some baby steps to 2.23 in 1967, and 2.36 in 1968 (Schmidt's last record). 1973 took us to 3.40 and 3.53, still for radio selected quasars, almost at once, then to 3.78 in 1982. The last time the most distant object was radio-loud was 1986, with a source at z = 3.80.

Radio-quiet quasi-stellars were bound to win out over radio loud ones, if only because there are so many more of them. The trick was to know where to look. And the answer came with objective prism or grism surveys. Meanwhile various other photoelectric detectors, culminating in CCDs, had begun to replace photographic emulsions as the spectroscopic materials of choice. Thus the lead in z-space was taken over by an optically-selected QSO at $z = 4.014$ in 1987, rising later the same year to 4.402 and to 4.733 in 1989. The current QSO record is $z = 4.897$, for PC 1247+3406, found in 1991, and the most distant radio-loud source is at $z = 4.46$.

Efforts to detect and recognize distant galaxies had, of course, gone on. Identification of 3C and other sources continued, largely under the leadership of Spinrad, and they reached $z = 1.27$ in 1982 and $z = 1.82$ for 3C 256 in 1985. Other groups, often using the Anglo-Australian Telescope, then got into the act, increasing the radio galaxy record to 2.3 in 1988 (actually a KPNO 4-m result) and $z = 3.80$ in 1990. This was followed by two medium-sized steps to 4.25 in 1994 and to 4.41 in 1997.

Ordinary galaxies are, of course, still commoner than either radio-emitting ones or active ones, and we expect this also to be the case at large redshift, though the galaxies or their precursors may not look like anything Hubble would have wanted to classify. The sky is, however, heavily smudged with faint, fuzzy things, and there will be a great many intrinsically faint, nearby galaxies for every bright, distant one. A few galaxies at redshifts near one had turned up in large surveys, and one quite recently beyond two. But, as usual, it was new ways of looking at things that made the difference. The first winner was to look for galaxies with strong Lyman alpha emission at the same places in the sky and with the same redshifts as Lyman-alpha absorbing clouds seen in the spectra of more distant QSOs. This was so obviously the right thing to do that the relative paucity of results led people to wonder whether an epoch of galaxy formation (when many spheroids made many of their stars in a short time) had ever actually happened (and the answer remains, "not entirely"). Searches for Lyman alpha emission through filters chosen to pick out a particular range of redshifts were similarly slow to yield results. But a handful of galaxies with strong Ly-alpha emission lines eventually turned up, with a current maximum redshift of 4.55 (for one not near a quasar).

Most recent and spectacularly successful is the technique called Lyman drop-out. The idea is that as 1216 A moves into a particular broad-band filter pass (U, B, V, etc.) the enormous forest of Lyman-alpha absorption lines eats most of the photons, and a galaxy becomes very faint in that band. Charles Steidel and his colleagues pioneered the method, looking first near distant QSOs, where they knew they would find something, then eventually at blank fields, including the Hubble Deep Field. Candidate galaxies must, of course, have their redshifts confirmed with spectra (required to show two lines or a line and a continuum break at a wavelength ratio recognizable as something you expect). But at the time the conference was being organized, the largest known z (apart from the cosmic background radiation!) belonged to a $z = 5.34$ galaxy found this way. It was 5.67 in October, with larger values definitely expected any day (at least in the windows where the lines can peek through the earth's OH and other atmospheric features - the next is at 6.7), and with several strong candidates for $z = 6$-7 in the HDF. Some are so faint that even

Keck and other 8-10m ground-based telescopes may not suffice to get redshifts, leaving us waiting for the NGST.

LOOKING AHEAD

In light of the rapid growth of maximum z and some theorists' predictions that we have already truly reached the era of first galaxy formation, the conference dinner speaker, Vera Rubin, suggested a modest prize, to be given to the first person to find an honest, spectroscopic redshift of 7.0 or larger, the prize to include an antarctic rock.

Of course, finding distant galaxies (etc.) is just the first step. Understanding is still more important, and we come back more or less to the conditions described by W.C.G. Whether in about 1920, from whose poem the title of this talk was taken:

Beyond the bright searchlights of science
Out of sight of the window of sense,
Old riddles still bid us defiance,
Old questions of Why and of Whence.

And there can be no better person to introduce you to those questions than the next speaker on the program, Sir Martin Rees.

Further details of the discovery of the cosmic distance scale and its changes can be found in PASP 108, 1073; in Space Science Reviews 79, 793, and in the 1997 proceedings of the STScI symposium on The Cosmic Distance Scale. It is reasonable to expect that fairly firm values of the cosmic distance scale and of the epoch at which the first galaxies, QSOs etc. formed will be established not too far into the 21st century.