Historical Development of Modern Cosmology
ASP Conference Series, Vol. 252, 2001
V.J. Martínez, V. Trimble, and M.J. Pons-Bordería

The Quest for the Edge of the Universe

Virginia Trimble

Astronomy Dept., Univ. of Maryland, College Park MD 20742 USA, and Physics Dept, Univ. of California, Irvine CA 92697 USA

Abstract. An overview of at least portions of the summer school is presented in the form of a discussion of how mankind's ideas of the size (and age) of the world in which we live have expanded, more or less monotonically over most of recorded human history, especially during the 500 year lifetime of the University of Valencia.

1. Introduction

There are many different stories to be told in connection with the development of our current cosmological ideas, and at least three ways of telling them. The ways include technological (adopted by Longair and new to most of us, though I confess he left out one of my personal heroes, James Bradley and his precision zenith sector, with which aberration of starlight was discovered), observational, and theoretical. Most of the speakers were indeed theorists, perhaps because the observers were on mountain tops observing, but Novikov, Einasto, Peebles, and Jones focussed on specific observations of the cosmic microwave background, dark matter, large scale structure, and so forth.

It also became clear during the week that some of the stories have both an eastern hemisphere and a western hemisphere version. Novikov and Einasto, for instance, described the recognition of various important aspects of the CMB and of dark matter where the scientific content was familiar to many participants but where the literature, published by colleagues from the former Soviet Union, was much less familiar.

More or less orthogonal to these ways of tracing history are the particular stories one wants to tell, or, in different language, the particular questions one might want to answer. These can include:

- What is the purpose of the universe? This has rather gone out of fashion, though the talks of Barrow and Heller perhaps had some bearing on it.

- What is our purpose in the universe? Ditto

- How big (or how old) is the Universe? In this context, infinity is not just a really big number, as Silk made clear.

- What is the universe made of? The answer undoubtedly includes baryons (addressed by Kragh), neutrinos (addressed by Valle), and other sorts of dark matter and early universe relics (addressed by van den Bergh, Einasto, Kolb, and others).

• What constitutes a legitimate method of studying the universe? In this context, the historical talks by Hoskin, Coles, Christianson (on Hubble with his peculiar aversion from theory), and Bonometto were particularly insightful, though the contrast between analytical and numerical methods (Icke and Yepes) is perhaps a hotter topic just now.

• Have we got hold of at least roughly the right end of the stick? Most of us tacitly assumed so (as has been the custom of scientists from at least the time of Aristotle to the present). Narlikar undoubtedly suggested the answer "no", and Bonometto may have done so at a more subtle level.

• What is the structure and evolution of the universe on the largest scales (including perhaps outside our own four-dimensional space–times)?

This last is, in some sense, a definition of cosmology, as perceived by most contemporary astronomers. Thus we all addressed some aspects of it, while leaving a number of other aspects for another gathering, when some other university reaches its 500th anniversary.

One neglected aspect was the specific values of the cosmological parameters. The following is a self-consistent set, derived from supernovae, fluctuations in the CMB, large scale structure, gravitational lensing and more traditional methods. Nearly the same numbers were suggested by several members of a panel discussion at the 1997 General Assembly of the International Astronomical Union in Kyoto and re-endorsed by speakers at the IAU in Manchester in August, 2000.

H	=	65± 10 Km/sec/Mpc
Ω_b	=	0.03
Ω(HDM)	=	0.02
Ω(CDM)	=	0.25 – 0.30
$\Omega(\Lambda)$	=	0.65 – 0.7, sometimes called dark energy
k	=	1, that is, flat space
n	=	1, the Harrison–Zeldovich spectrum of initial perturbations

but the normalization of that spectrum remains something that we are unable to calculate or predict and must be tied to the quadrupole moment measured from the COBE satellite.

There are, of course, other questions that must be left for the future, including, What is the dark matter? How did galaxies and large scale structures and streaming arise without ruffling the CMB? What/when came before the Big Bang? Are there other universes? It currently seems very probable that some or all of these questions, or at least their answers, will turn out to be closely related, as hinted in Kolb's talk. I will focus here on the "how big" question, and on the observational track.

2. Sizes and Distances of Astronomical Objects: The Beginnings

The specific objects and systems addressed will be the earth, earth–moon, earth–sun, stars, the Milky Way, and the Universe, and the time frames in which the questions were first meaningfully asked and later answered will be crudely divided into mythic (not always before the first written records, but before the first written records of a particular culture), early literate (from the Greeks to the middle ages), renaissancetto 18th century, 19th century, 20th century, and the future. In case I forget to mention it later, in only one case (the earth–moon distance) was the size question both asked and answered in a single one of these periods.

Once upon a time, earth, world, and universe were all more or less the same concept. In this context, myths that have come down to us (in Genesis and its Babylonian ancesters, from the Egyptians, Native Americans, and so forth) all seem to envisage a size of the world that humans might traverse in a generation or a god in a day. Apollo and his chariot come immediately to mind. The Egyptian sun god Aten changed from a day boat to a night boat each sundown to start down around under the earth, and back to his day boat each morning. It is entirely my own theory, without any supporting evidence, that part of the motivation for this picture was to account for the apparent slight slowing of the motion of the sun near the horizons, which we now attribute to refraction in the earth's atmosphere. And the height of the curved heavens (the arched body of the goddess Nut in the Egyptian version) was comparable with the diameter of earth.

Many myths (again Genesis, the Maori tales of which ancestral families came in which boats) imply a total time scale from some sort of divine creation process to the present measurable in human generations, perhaps 10^3–10^4 years. A clear exception is an Indian tradition, implying time scales of (at least) billions of years. A wide disparity in age estimates persisted down to relatively recent times. Numbers entertained near the start of the 20th century ranged from not much more than 10^7 (the Kelvin–Helmholtz time for the contraction of the sun, so called because it was discovered by Mayer and Waterston) to 10^{12-13} years, deduced by James Jeans from his understanding of how binary pairs of stars and clusters of stars and galaxies have evolved dynamically, one encounter at a time.

3. From Earth to Moon to Sun

Introductory astronomy textbooks tell us that the Greek-cultured but Egyptian-resident Eratosthenes (who lived from -276 to -195, backward, as people did in those days) observed that, at noon on 21 June, a stick inserted vertically into the ground near modern Aswan casts no shadow, while one at Alexandria casts a shadow 12% of its own length. In other words, the circumference of the earth must bear the same ratio to the Aswan–Alexandria arc distance that 7° does to a whole circle. His answer for the circumference was 250,000 Roman stadia, and no, we do not know the length of a Roman stadium accurately enough to be sure just how well the surveying had been done. This measurement must necessarily represent the result either of an early international collaboration or of an early multi-year grant.

The chinese equivalent is embodied in a god or hero named Pan-ku, who separated the heavens and earth by standing in between and growing 18,000 feet per day for 10 years. The resulting 12,400 miles (assuming "feet" were the same size then as now) is close enough to an earth diameter to suggest that some similar sort of measurement had been made, or at least heard of, there.

Once you knou the size of the earth, the distance to the moon is easily determined from the angular diameter of the earth seen by an observer on the moon. At least this is conceptually easy, though Aristarchus of Samos (-310 to -230), who thought of it, was forced to use as a proxy the curve of the earth's shadow as it crossed the moon during a lunar eclipse and the duration of the eclipse. His answer of 60 earth radii for the earth–moon distance is good enough for many purposes.

Aristarchus also supposed that one might work geometrically outward to the earth–sun distance from the unequal duration of the phases of the moon. This would be OK if the moon's orbit were a perfect circle, though the difference would then be too small for his contemporaries to have measured it. In fact, they were fooled by the non-circular lunar orbit (responsible for most of the day-or-so inequality) and wrongly calculated an earth–sun distance of about 1000 earth radii.

Numbers of this general sort endured through the transfer of astronomical knowledge from the last Greeks to the Arabs and back to Europe in the High Middle Ages, and Copernicanism had largely triumphed before a more promising method than Kepler's nested platonic solids could be brought to bear on the problem.

And then, one is inclined to say, there were the Middle Ages. This is actually almost true of the history of astronomy; and the end is marked by the work of Thomas Aquinas (1225–74), whose synthesis of Catholic church doctrine with Aristotelian phylosophy incorporated an earth-centered cosmos, in which celestial objects were perfect, spherical, unchanging, and made of a fifth element, light traveled with infinite speed, and many other items the disproof of which can be counted as steps forward toward modern science.

In a Copernican solar system with Kepler's laws, once you have measured any distance, geometry gives you the rest. And, in 1672, Mars came unusually close to earth at opposition, permitting a trial of geocentric parallax measurement (which Tycho had already used to confirm the Greek lunar distance). Cassini (of the Saturnian ring division) and, we must deduce, a colleague, attempted simultaneous measurements of the position of Mars as seen from Cayenne (of the pepper) and Paris (of the plaster) while Flamsteed (who perhaps also saw Cas A as a supernova in 1680) and Cassini and Roemer (who first measured the speed of light) had a go at earth rotation parallax. They got $19-23''$ as the Martian parallax, corresponding to $10-12''$ for the sun. Newton apparently used an even less good value in an early version of *Principia*, putting the sun at a distance of 20,000 earth radii.

A century later, careful timing of the 1769 transits of Venus settled on $8.4-8.8''$. This is still not quite as good as you might expect it to be, because the main source of error was not clock errors but the difficulty in deciding just when the black dot crosses the solar limb (it sort of seems to bleed in and then bud off). Geocentric parallaxes of earth-crossing asteroids had refined the number further before radar reflections and interplanetary probes determined the length of the Astronomical Unit to umpteen significant figures.

4. On to the Stars

The heavens from mythic times at least to Anaximander (-610 to -545), for whom the sun was more distant than any of the planets or stars, were not much bigger than earth. From Aristotle (roughly) to Kepler, stellar distances were generally thought to be 10 or 20 times the earth–sun distance. Kepler attempted to make the numbers more precise by requiring the spheres of the several planets to nest around and between the five platonic solids (ordered so as to give roughly the right ratios of separations and period implied by observations and his $P^2 = a^3$ law). He obtained a somewhat larger radius for the sphere of the stars.

But Tycho had set an upper limit of $25''$ to what we would now call heliocentric parallax, which, in a Copernican universe, forced the stars out to at least 3700 earth–sun distances. Notice that he just missed the $20''$ aberration of starlight! Meanwhile, over several centuries, the idea that "stars are suns" (but the sky is nevertheless dark at night) occurred to Nicolas of Cusa (1401–1464, who also thought they would share the composition of the earth and pointed out that an infinite universe has no center), to Thomas Digges of England (who published in 1576), to Giordano Bruno (before his 1600 burning for various heresies and, it would seem, a considerable degree of political ineptness), and to others, again implying stellar distances of thousands of times that to the sun.

As Copernicanism and the ordinariness of the sun gradually became normative ideas, the search for parallax continued. It led Bradley to his 1729 discovery of aberration and William Herschel (who had already used the "stars are suns" idea to trace out a size and structure for the Milky Way that dominated for more than a century) to his discovery of the orbital motion of binary stars and thus to a confirmation of the physical association of the pairs, first deduced a generation earlier by John Michell, on the basis that there were too many close pairs for them to be chance alignments. Herschel had focussed on close pairs of stars with different apparent brightness precisely because he thought that relative parallax would be most obvious there if all stars had the same real brightnesses.

Three astronomers, working independently, found statistically significant parallaxes for three different stars, selected to have large proper motion, small apparent magnitude, or both, in 1836. The winners were F. Bessell (parallax = $0.29''$ for 62 Cygni whose record proper motion is $5.2\ ''$/yr), F. Struve ($0.12''$ for Vega, with $\mu = 0.35\ ''$/yr, but, of course, a good deal brighter), and Th. Henderson, working from the Cape of Good Hope ($0.75''$ for Alpha Centauri, $\mu = 3.7\ ''$/yr and still the closest known). Longair notes (elsewhere in these proceedings) that Struve and Bessell were using new, improved technology by the same designer.

Most naked eye stars are intrinsically a good deal brighter than our sun. Thus Herschel's assumption of equality led to errors in distance of a factor 3 for Procyon, 5 for Sirius, 10 for Arcturus, 30 for Canopus, and 100 for Rigel. This, in combination with his complete neglect of the possibility of interstellar loss of light (scattering or absorption) resulted inevitably in his tracing out a Milky Way far smaller and more nearly sun-centered than the real one.

5. Milky Way Issues

Early in the 20th century, the three main galactic disputes concerned the total size of the Milky Way, our galactocentric distance, and the uniqueness of the Galaxy, with the "island universe" hypothesis as the alternative. From William Herschel to Kapteyn and on to Heber Doust Curtis, the counting of stars vs. spectral types or colors, with calibrations from secular and statistical parallax, resulted in typical diameters of 3–17 kpc and a galactocentric distance of nearly zero. The primary cause was total neglect of interstellar absorption and scattering of starlight. Though Kepler had been among those supposing that light would be impeded on its way to us (he had himself singularly poor eyesight), his view became the dominant one only after Robert Trumpler showed, in 1930, that a plot of angular diameter vs. apparent brightness for open star clusters made sense only if there was about 1 magnitude per kiloparsec of absorption in the Galactic plane. Though Mineur (1946) recalibrated the period–luminosity relation of the Cepheid variables allowing for such extinction, most of the rest of the astronomical world did not catch onto its full significance until after the 1952 Rome General Assembly of the IAU, when Baade announced his failure to resolve RR Lyrae stars in M31 and Thackeray reported success in the Magellanic Clouds. Trumpler himself attempted to "agree with all previous speakers" by drawing a Milky Way centered where Shapley had pointed, but with a "Kapteyn universe" off to one side and centered at the sun. The residual milk of truth in this is the existence of a number of bright OB stars in what is also called Gould's Belt.

Shapley, who generally features as a culture hero in this territory, made the same mistake of neglecting absorption in his calibration of the distance scales for Cepheid and RR Lyrae variables in the field and globular clusters. The effects were rather subtle, since his field calibrators were in the disk and absorbed, while his cluster sample was at high galactic latitude and largely unobscured. One result was the illusion that a single period–luminosity relation extended over all such stars. The other (once the relation was calibrated with statistical parallax for a handful of field Cepheids) was to put the globular clusters and galactic center considerably further from us than the best modern estimates.

A snapshot at the 1920 Curtis–Shapley debate finds Curtis advocating a 3 kpc Milky Way, with the sun very near the center and other, comparable galaxies well outside, while Shapley insisted upon a galactocentric distance up to 20 kpc, a total galactic diameter of 60 kpc or more, and no external galaxies. Their evidence and arguments have been rediscussed many times. My version is Trimble (1995).

Since that time, R_o has bounced between about 10 kpc (Oort in the 1920's and an IAU vote at the 1964 Hamburg General Assembly) and 8.5 kpc (Baade in the 1950's and a 1984 IAU vote in Delhi, India) The average of values published in the last five years is probably closer to 8 kpc, or even a bit less, but the total extent (at least of the dark matter halo) implied by orbits of globular clusters and dwarf galaxy satellites is something like 100 kpc, large enough to satisfy even Harlow Shapley. By-the-by, on Shapley's scale, some of the more distant globular clusters were 30–40 pc in size, and the implied masses would have been sobering, if he had known their velocity dispersions.

First attribution of spiral structure to the Milky Way has generally been credited to Cornelius Easton (1900), but there was a "prediscovery" by Stephen Alexander (1852), professor of mathematics and astronomy at the College of New Jersey (a.k.a. Princeton), to which I was guided by a paper on spiral structure by George C. McVittie and Cecilia Payne–Gaposchkin. Easton wins the prize for creative artwork, however. His drawing shows a badly squashed spiral, centered off to one side in the direction of Cygnus, while the circle representing the total diameter of the galaxy is solemny centered on the sun, which is wrapped in a single, loosely wound, outer arm.

And, of course, other galaxies do exist, though it became clear that many of them are as large as or larger than the Milky Way only after the multiple revisions of the extragalactic distance scale in our century. Early in the process, Jeans had said that, if the others were islands, then the Milky Way was a continent.

6. Existence of other Galaxies and the Extragalactic Distance Scale

Between 1917 and 1921, Curtis, Shapley, and Knut Lundmark all considered the possibility that there might be two classes of novae. If so, and if all the novae seen in M31 except the 1885 event were of the fainter class, then the nebula must be 200–250 kpc from us, a plausible answer in the eyes of Lundmark and Curtis, though regarded as ridiculous by Shapley, who took S And to be a typical nova and put the nebula inside his larger Milky Way. Ernst Opik pushed the distance out to 450 kpc in 1922 by assuming that the nebula should have the same ratio of mass to light as the solar neighborhood. (Remember that V.M. Slipher had recorded its spectrum as early as 1912, so Opik knew its absorption lines were tilted and the nebula was rotating).

Edwin Hubble settled the existence issue once and for all by about 1925, with his report of Cepheid variables in NGC 6822 and M31 (with a possible pre-discovery by Humason about 1920). Hubble was not the first to look for a distance-dependent K term in an analysis of line shifts of nebulae, nor even the first to find one (see Trimble 1996,1997a for many of the primary and some secondary references to work in this period), but he was the first to be widely believed.

Hubble fitted his K-parameter (later called the Hubble constant, H) in several different ways, finding values between 500 and 550 km/sec/Mpc and concluding that the uncertainty was about 10%. This 10% uncertainty has been the one invariant with time in this branch of astronomy and applies also to the number reported at the August, 2000 General Assembly of the IAU by the Hubble Space Telescope Key Project Team (74 km/sec/Mpc). Hubble (1929) is, therefore, generally described as having "discovered the expansion of the universe", although his own interpretation of the velocity–distance relation, along with galaxy number counts and brightnesses, veered between true expansion and something like tired light several times. It is possible to argue that this issue was not finally resolved until the demonstration in the 1990's that light curves of distant supernovae are not only redshifted but time dilated.

In the 70+ years since Hubble's key paper, two concurrent processes have continued to expand our horizons (see Trimble 1999). These are the measurements of ever larger redshifts and the decrease in the best-bet numerical values of Hubble's constant. Both look step-like in plots of numbers vs. time. The possibility of reaching to ever larger redshifts has been provided largely by technological advances: larger telescopes (from 36″ to 100″ to 200″ and beyond), better detectors (from photographic emulsions of gradually increasing sensitivity to the scanners pioneered by Spinrad and Gunn from the mid 1970's, to CCDs and soon, perhaps, superconducting tunnel junctions), not to mention the addition of radio and X-ray surveys to our toolkit for picking out unusual objects.

Hubble's 1929 diagram reached clear out to 2 Mpc and 1000 km/sec (i.e. z=0.003), using spectrograms exposed by V.M. Slipher at Lowell Observatory. The maximum redshift was 0.2 by the time Milton Humason hung up his observing boots in the early 1960's. At last look, galaxies, quasars, and gamma ray bursters had all been spotted to z=4 or more, and the barrier near z=6 set by useful spectral features disappearing into the infrared sky background are likely to fall soon to confirming infrared spectra of objects chosen from their colors (especially Lyman-limit drop-outs) in SDSS, 2MASS, DENIS, and other surveys.

In complimentary fashion, many of the changes in estimates of the Hubble constant have been driven by thinking harder about the meaning of data already in hand, including interstellar absorption, single stars vs. compact clusters, and, particularly, proper accounting for Malmquist bias and the Scott effect. These are sometimes confused, but the former means that, in distant samples, you will be able to catch only the brightest part of a distribution of intrinsic luminosities (even if these are the same at all distances), while the latter means that, in distant samples, because volume is proportional to d^3, you may find very rare kinds of objects, brighter than any in your local sample. If there are three cosmologists in a room, it is very probable that two will say that the third does not understand Malmquist bias (see Trimble 1997b).

Despite this, modern cosmology has expanded from what Sandage once described as "the search for two numbers" (meaning H and the deceleration parameter q) to encompass physics of the early universe, precision measurements of a dozen or so significant numbers (including separate densities in baryons, hot and cold dark matter and "other"), galaxy formation, and much else discussed at this school and elsewhere. I would not, however, bet anything more valuable than bottles of good Spanish wine on a particular value of H more tightly constrained than 65±20 km/sec/Mpc, or on a cosmic age more specific than 10–20 Gyr.

Acknowledgments. It is the traditional prerogative of the last speaker at a conference to extend the collective thanks of the participants to those who have made the gathering possible. Our very great gratitude goes to Vicent Martínez, chair, and María Jesús Pons Bordería, secretary of the conference, and to Alicia Gómez, who helped enormously with arrangements in the lecture hall. Equal appreciation, of course, belongs to the organizations who provided financial sup-

port and so made it possible for both lecturers and students coming from places as far away as Princeton and Pune to participate. These are Cinc Segles de la Universitat de València, Universidad Internacional Menéndez Pelayo, Caja Rural de Valencia, Generalitat Valenciana, Caja de Ahorros del Mediterráneo, Ciutat de les Arts i de les Ciències, the British Council, and NASA. From these names, you may gather, first, that we had the support of governments, commerce (the Cajas are banks), and education, and second that both Spanish and Catalan are widely used in Valencia. That most of the lecturers and many of the students knew not a word of either did not in any way interfere with our feeling at home in the city.

References

Alexander, S. 1852, AJ, 2, 95 & 105

Easton, C. 1900, ApJ, 12, 136

Hubble, E.P. 1929, PNAS, 15, 168

Mineur, H. 1946, Ann. D'Ap., 7, 160

Trimble, V. 1995, PASP, 107, 1133

Trimble, V. 1996, PASP, 108, 1073

Trimble, V. 1997a, Space Sci. Rev., 79, 793

Trimble, V. 1997b, in The Extragalactic Distance Scale, ed. M. Livio et al. (Cambridge: CUP), 313

Trimble, V. 1999, in AIP Conf. Ser. Vol. 470, After the Dark Ages, ed. S.S. Holt & E.M. Smith, 3