Title
H0: The incredible shrinking constant 1925-1975

Permalink
https://escholarship.org/uc/item/0tg0q2qx

Journal

ISSN
0004-6280

Author
Trimble, V

Publication Date
1996-12-01

DOI
10.1086/133837

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed
1. THEORETICAL UNDERPINNINGS

All discussions of modern cosmology begin with Einstein’s 1916 publication of the theory of general relativity, and it must be said right off the bat that if GR, or at least something with essentially the same classical limit, is not the right theory of gravitation, then all bets are off, and we can all go home. His own exploration of “cosmological implications of the general relativity theory” appeared in 1917 (Einstein 1917) and, notoriously, included the cosmological constant needed to make space–time static. That Einstein later regretted the addition is not actually relevant to whether the constant is or is not zero. His universe had a uniform density of matter, uniquely related to the value of the cosmological constant and a characteristic size scale.

Willem de Sitter (1872–1935), motivated perhaps by Machian considerations of the nature of inertia, identified another solution of the Einstein equations that is also static, contains a cosmological constant, and is completely void of matter or radiation. His relevant publications were “On Einstein’s theory of gravitation and its astronomical consequences” (de Sitter 1916) and “On the curvature of space” (de Sitter 1917). Space and time coordinates intermingle in his metric (which is the equivalent of four-space embedded in Euclidean five-dimensional space). As a result, photons emitted by moving test particles show Doppler effects (not always of the same sign as the test particle velocity). Redshifts should predominate and scale roughly as the square of the distance to the emitting particle. This was the only sort of non-static behavior generally known in the scientific community before about 1930. As a result, most of the pre-Edwinian efforts to correlate nebular redshifts with distances started out by looking for quadratic relationships. It can be shown that the Einstein, de Sitter, and special relativistic line elements exhaust the possibilities for static solutions (Friedmann 1922, of whom more in a moment; Tolman 1929a) and even for stationary solutions (Robertson 1929).

Meanwhile, largely unnoticed, and illustrating the principle that even back then “nobody reads the literature,” Alexander Friedmann (1888–1925) had published “on the curvature of space” and “on the possibility of a universe with constant negative curvature” (Friedmann 1922, 1924). The former is available in English in Lang and Gingerich (1979), but all translations shown here in quotes are my own interpretation of what the words would have meant to a contemporary monoglot astronomer. Cognoscenti will already have recognized that Friedmann was describing what we now think of as critical and open models of the universe (that is, ever-expanding ones with zero and negative curvature, respectively).

Soon after, Abbe George Lemaître (1894–1966) provided a “note on de Sitter’s universe” (Lemaître 1925) and went on to propose “a homogeneous universe of constant mass and croissants made of rayon,” oops, sorry, “increasing radius” (Lemaître 1927, with a post-Hubble English version called “the expanding universe,” Lemaître 1931a).

The year 1928 saw Edwin Powell Hubble elected to the presidency of the International Astronomical Union Commission on Nebulae at the third general assembly in Rome in succession to Vesto Melvin Slipher (of whom more shortly), the second president.

Hubble’s exposure to discussions of the de Sitter universe at the 1928 meeting was an important driver in his examination of redshift–distance correlations over the next year or two, according to Osterbrock’s (1993) article on Hubble’s cosmology (also the source of several other items mentioned here). Incidentally, Hubble always preferred to speak of nebulae; galaxies was Shapley’s choice.

Lemaître (1927) had shown that the Einstein static universe was unstable, a small kick sending it into never-ending expansion (or contraction). Thus only de Sitter-, Friedmann-, or Lemaître-type cosmologies could describe the real universe. And the decision amongst them would necessarily have to be made on observational grounds. Lemaître had also
combined early Slipher velocities with some sort of distance indicator of his own to suggest a constant of about 600 km s\(^{-1}\) Mpc\(^{-1}\) in a linear distance–redshift relation.

Arthur Stanley Eddington seems to have played a key role in introducing expanding models into mainstream astronomy, in discussions at meetings of the Royal Astronomical Society in 1930, in a paper in their journal ("on the instability of Einstein's spherical world," Eddington 1930), and in a 1933 book called *The Expanding Universe* (Eddington 1933, with later editions and reprints still to be found in most large university libraries). Lemaître (1931b) waded back into the fray with "a homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae." And Friedmann was dead, at age 37, in 1925, frustrating the intention of the young George Gamow to study with him in Petrograd, thus forcing Gamow into nuclear physics, with far-ranging consequences for modern cosmology!

William H. McCrea and George C. McVittie (1930) were among other early proponents of a universe that has expanded for all or most of its observable history, as was Hermann Weyl (1930). Howard Percy Robertson (1928, "On relativistic cosmology," 1935) and Arthur Gordon Walker (1935a, "On Riemannian spaces with spherical symmetry about a line and the conditions of isotropy in general relativity") put the metrics used by de Sitter, Friedmann, and Lemaître into the most general possible form, and the one that is still in general use. Another small sidelight: Robertson died in August, 1961, just as Jim Gunn was arriving to work with him at Caltech, changing the emphasis of his intended research, but somewhat less than Friedmann’s death reaimed Gamow.

2. REJECTING THE ALTERNATIVES

How sure are we that we live in a relativistic, expanding universe of finite age, so that measuring Hubble’s constant is genuinely an interesting thing to do? The main alternatives with reasonably long lifetimes have been tired light, expansion into previously existing space, and steady state. Hubble himself did not entirely reject the first possibility, believing that the matter could and should be settled observationally (Hubble and Tolman 1935, “Two methods of investigating the nature of the nebular red-shift”).

Even as Hubble’s classic paper was appearing on library shelves in 1929, Fritz Zwicky (1898–1974) proposed that photons might simply lose energy and become redshifted by virtue of their long journeys ("On the red shift of spectral lines through interstellar space," Zwicky 1929). This is generally called tired light. There are both theoretical and observational reasons for rejecting it. First, we think we have a good theory of the behavior of photons (quantum electrodynamics), and the Feynman graphs for tired light sum to zero.

Fig. 1—Published values of the Hubble constant from Lemaître (1927) to the hardening of the battle lines. Rectangle dimensions are intended to suggest a range of values/uncertainties or a range of dates. Except where the errors listed below are larger, all uncertainties were claimed to be of order 10% or less (occasionally much less). A straight-line fit to the numbers from 1927 to 1965 or so would have suggested that the Hubble constant might have become negative within a decade or two (discovered by astronomy graduate students at Caltech in the 1960s and undoubtedly by many others). This did not actually happen. The numerical values represented are Lemaître 600, Hubble 465, 513, 535; Hubble and Humason 526; Mineur 320; Behr 240; Baade and Thackeray 280±30; Hubble, Mayall, and Sandage 180±20; Sandage 75 (+75, −40); Holmberg 134±6; McVittie 143±27; Sersic 125±5; van den Bergh 100 (+20, −12), 120 (+25, −20); Ambartsumyan 70–100; de Vaucouleurs 125, 100±10, 100±10; van den Bergh 95 (+15, −12); Sandage and Tammann 45–60.
First of these is the Hubble–Tolman (1935) surface brightness test. They pointed out that the energy we receive (in ergs cm$^{-2}$ s$^{-1}$ sr$^{-1}$, for instance) will scale as $(1+z)^{-1}$ in a tired light universe and $(1+z)^{-4}$ in a relativistic, expanding one. I won’t attempt to explain why (in accordance with Ehrenfest’s theorem that it is quite difficult to explain something even if you understand it, and almost impossible if you don’t). If you are feeling brave, have a go at Sandage’s (1988) Annual Reviews article, which explains surface brightnesses and a number of other important cosmological measurables. Sandage (1992, another source from which I have cribbed heavily) is of the opinion that the surface brightness test is definitive and the $(1+z)^{-4}$ dependence seen. Other measurers of galaxian surface brightnesses are less sure (Djorgovski 1995). Incidentally, the other Hubble–Tolman (1935) test involves counting galaxies as a function of redshift and apparent brightness. It has been declared moot a countably infinite number of times, but note that the recent release of the Hubble Deep Field Image has triggered another round of counting.

A second more informative test can also be derived from the work of Richard Chase Tolman (1934; this book is the only one I bought in graduate school that wasn’t required. It cost $10 and is the source of some of the technical descriptions in Secs. 1 and 2 here). The test consists simply in recognizing that blackbody radiation in an expanding universe will cool with the expansion. This effect has arguably also been seen (Songaila et al. 1994).

The third test is, to my mind, the most persuasive one. In an expanding universe, events seen from far away will be time dilated by a factor $(1+z)$, while in a tired-light universe they will merely look faint. The absence of a correlation between time scale and redshift in the variability of quasars (etc.) has been one of the props of the “non-cosmological redshift” school. But we now have better distant clocks. Fritz Zwicky (1939) and Olin C. Wilson (1939) both noted that supernovae, being very bright and rather homogeneous, might be good distance indicators (van den Bergh and Tammann will tell you whether they were right about that point). Wilson in addition remarked that time dilation of their light curves should be detectable at large redshift, if the expansion were real. This has now been seen (Perlmutter et al. 1995, and an unpublished data base of about two dozen events with $z = 0.3–0.5$ studied by the same group).

The second alternative to relativistic expansion of space—time itself is motion of galaxies into previously existing (probably Euclidean) space from some center. If you wait a while and the galaxies initially had a wide range of speeds, the effect will be an isotropic linear velocity—distance relation, as long as you observe from somewhere near the center. This was the cosmology put forward by Edward Arthur Milne (1896–1950) under the title “A Newtonian expanding universe” (Milne 1934, 1935). In the same time frame, McCrea and Milne (1934) and Arthur Gordon Walker (1935b, and yes it’s the same one) wrote on “Newtonian universes and the curvature of space” and “On the formal comparison of Milne’s kinematical system with the systems of general relativity.” Both papers showed that there are close Newtonian analogs to the GR models. McCrea (1990) has provided a modern perspective on the Milne cosmology. The primary reason for rejecting this sort of picture is that we see absolutely no evidence for edge effects and, with the present very small values for anisotropy of the microwave radiation background, would have to be most Copernicanly near the center of the expansion. Milne’s ideas became more complex with time (e.g., Milne 1948) and eventually involved two separate time scales for electromagnetic and gravitational processes. This doubling provided one way out of the 1940s discrepancy between the age of a relativistically expanding universe (about 2 Gyr) and the age of the oldest earth rocks (3–4 Gyr). Walker, as well as McCrea, was Milne’s student, and Robertson was also somewhat influenced by him.

Mention of the age discrepancy brings us naturally to the third alternative, steady-state cosmology, in which new matter appears at just such a rate as to keep the average density of the expanding universe constant. Hermann Bondi and Thomas Gold (1948, “The steady state theory of the expanding universe”) and Fred Hoyle (1948, “A new model for the expanding universe”) had among their motivations the need for a longer cosmic time scale than the reciprocal of the contemporaneous value of $H$ (about 500 km s$^{-1}$ Mpc$^{-1}$, $1/H$ about 2 Gyr). It is, I think, a fair statement that lots of astronomers, including the present writer, took the steady-state alternative seriously between 1948 and somewhere around 1955–65. The number who still take it seriously is now very small. Brush (1993) provides one view of how this happened.

Notice that the redshift–distance or Hubble relation has quite different meanings in these alternatives. In a tired-light universe it would be well defined and, presumably, a clue to understanding quantum field theory. In an exploding universe it is both meaningful and an indicator of the time since the explosion, while the steady-state value of $H$ tells us nothing about the age of the universe as a whole, but only about the average age of galaxies to be found in some representative volume. In contrast, $H$ is not even well defined, let alone meaningful, if Arp’s (1989, e.g.) non-cosmological redshifts dominate.

I have described all these alternatives as “rejected,” but honesty compels the admission that there are still adherents of tired light models (e.g., many of the participants in the conference whose proceedings were edited by Peratt 1995) and also of some modified form of steady state (primarily Sir Fred Hoyle and people who have worked closely with him for long periods). Finally, Segal (1972; Segal and Nicoll 1996) continue to find a quadratic, in their words “Lundmark” rather than “Hubble,” relation between redshift and distance. The underlying theoretical ideas are apparently very old (Robb 1936), and I do not profess to understand them. The quadratic relation can actually be ruled out on observational grounds, at least if real motion of any kind is involved (as shown in Otto Heckmann’s 1942 volume, Theorien der Kosmologie, untranslated to this day, though I made a crude attempt long ago as a way of fulfilling...

3. THE PREHISTORIC PERIOD—MOSTLY OBSERVATIONS

Vesto Melvin Slipher (1875–1969) measured the very first Doppler shift for a spiral nebula—and also all of the next few dozen. His first was M31 (the Andromeda nebula) in a 14-hr exposure with the Lowell 24-inch Clark refractor on 3–4 December 1912. He reported a negative velocity near 300 km s\(^{-1}\), the largest then ever seen. Notice that this is genuinely a Doppler velocity, equal to the vector sum of the rotation of the Milky Way and its mutual orbit with M31. From now on, most of the spectral shifts mentioned will be cosmological redshifts, caused by the expansion of the universe, which are not the same as Doppler shifts (caused by motion through space). Slipher’s (1913) first M31 spectrogram is reproduced in Vol. II of Russell et al. (1927), and one can only say that one hopes the absorption lines were clearer on the original plate! By the time of the August 1914 AAS meeting in Evanston, IL, Slipher had 15 velocities, most of them positive.

Percival Lowell had instructed Slipher to study M31 on the assumption that it would provide insights into the formation of our own solar system. But, by 1917, when Slipher had accumulated 25 spiral velocities, he was of the opinion that the data provided some support for the island universe theory. His velocities ranged from −300 to +1100 km s\(^{-1}\), with positive values outnumbering negative ones 21 to 4 (Slipher 1917, “A spectrographic investigation of spiral nebulae”). His eventual “personal best” was +1800 km s\(^{-1}\) for NGC 584. Fainter, more distant, more redshifted galaxies exceeded the reach of the 24-inch telescope, even with Slipher’s improved spectrograph. Incidentally, Slipher’s brother, Earl C. Slipher, was also an astronomer, working at Lowell Observatory at the same time. He recorded many excellent images of Jupiter, Saturn, and their satellites.

Milton Lasell Humason (1891–1972) was the next major collector of redshifts, beginning in about 1927 (Humason 1927, “Radial velocities in two nebulae”). He soon more than doubled Slipher’s record redshift with +3779 km s\(^{-1}\) in NGC 7619 (Humason 1929), and doubled it again to +7800 km s\(^{-1}\) for NGC 7619 in the Coma cluster (Humason and Pease 1929). An improved spectrograph on the 100-inch telescope passed the 15,000 km s\(^{-1}\) mark (ζ = 0.05; Humason 1931, 1934, 1936). Humason (unlike Hubble) lived to use the 200-inch with some regularity, retiring in 1957 with a “personal best” of \(v = 60,000\) km s\(^{-1}\) for galaxies in the Hydra cluster. Many of the largest values were achieved with multi-night exposures, and Humason felt that the 200-inch could not reach beyond \(z = 0.2\), at least with detectors of the era. He was responsible for the montage of spectrograms and galaxy images, using apparent diameter as a distance indicator and illustrating Hubble’s law pictorially, that still appears in a large fraction of introductory astronomy texts. Any mention of Humason is considered incomplete if it does not say that his first involvement with Mt. Wilson Observatory was as a mule driver on the then-unmotorable road. For some reason, his 1950 D.Sc. from Lund University is less often highlighted.

Until about two years ago, I had always supposed that all or most of the velocities used in Hubble’s 1929 paper came from plates he had himself exposed. This is utterly false. It is also not absolutely certain that Hubble was the first to spot Cepheids in M31. Humason told Sandage in 1956 (and stood by it to his death; Christenson 1995) that he had found them earlier and marked them on some plates of M31, when he and Harlow Shapley were both connected with Mt. Wilson, but that Shapley “then calmly took out his handkerchief, turned the plates over, and wiped them clean of Humason’s marks.” It is at least possible chronologically. Humason was promoted to night assistant (from janitor) on the 100-inch telescope in 1919, about two years before Shapley moved on to Harvard.

Two important semi-theoretical considerations belong to this period. First, Ernst Julius Opik used the first of what one might call “ratio methods” of distance determination (Opik 1922), that is, methods that depend on your knowing the correct numerical value for the ratio of two physical quantities that depend on different powers of distance. He assumed that the Andromeda Nebula had the same luminosity-to-mass ratio as the solar neighborhood (0.38 in solar units) and took the rotation speed of 157 km s\(^{-1}\) implied by Slipher’s spectra to deduce a distance of 450 kpc. This was roughly double the distances found between 1917 and 1919 by Heber Doust Curtis, Harlow Shapley, and Knut Lundmark (1889–1958) using novae and single bright stars as standard candles (Curtis 1917; Shapley 1917; Lundmark 1919).

Second, Gunnar Malmquist (1893–1982) enunciated his bias. The idea is a simple one (Malmquist 1920, 1924). If you are studying a class of objects whose real brightnesses vary over some range, then you will see examples from the full range among nearby objects, but only the brightest ones far away. And, if you fail to allow for this, you will think the more distant ones are closer than they really are, because they look so bright. Proper correction for the bias requires knowing the actual distribution of brightnesses of the distant objects. Taking this a priori to be the same as the distribution for the nearby ones is an assumption that cannot always be justified. Any large gathering of observational cosmologists today will include at least one person who thinks that someone else in the room does not understand the Malmquist effect.

As Slipher’s (mostly) redshifts became known through the community and came to be perceived as real velocities, a number of astronomers attempted to correlate them with positions on the sky and distances to the nebulae. Some were motivated by early interpretations of the de Sitter universe and some by the more traditional problem of finding the solar motion and \(K\) correction relative to any astronomical population you could think of.

No two secondary sources give the same list of Hubble forerunners or the same interpretation of what each author thought he had shown. Going back to the original papers does not help quite as much as you might think. The list that follows is roughly chronological, from 1916 to 1928, and most of the original references can be found in Smith (1982):
O. Truman, R. Young and W. Harper (solar motion only, 1916); G. Paddock (the addition of a distance-independent $K$ or redshift term, 1916); Carl Wirtz and Knut Lundmark (1918–21 solar motion with $K$ terms); Gustav Strömbäck, Ludvik Silberstein, and A. Dose (1924–26 with $K$ still a constant); Lundmark (1925, allowing for the first time for a distance-dependent $K$ term, both linear and quadratic); Howard Percy Robertson, Georges Lemaître, and by implication Humason (or perhaps Hubble speaking through Humason’s pen; see Osterbrock 1993) between 1926 and 1928.

At least a few of these explicitly mention or imply numbers for what we now call the Hubble constant. Duerbeck and Seitter (1996) have provided the most recent discussion of what those numbers really are. Among the more clearly defined are 625 km s$^{-1}$ from Lemaître, 460 from Robertson, and (immediately post-Hubble) 465 from de Sitter and 432 and 290 km s$^{-1}$ from Jan Oort (1931, not cited by Smith).

Several of these papers appear distinctly odd to modern eyes, especially Silberstein, who attempted to include the globular clusters by allowing velocities to be either positive or negative and looking for correlations between distance and absolute value of velocity (speed, that is!). Apparently they were pretty odd even at the time, since Hubble seems to have gone out of his way to avoid seeming to indulge in even very mild theoretical speculation.

Lundmark was alone in including a quadratic term. Not surprisingly, he found one; including another variable parameter always improves the fit to your data sample. The quadratic term was negative and seemingly small; but it would have limited observable velocities to about 3000 km s$^{-1}$. The de Sitter model (at least in some interpretations) predicts a positive quadratic term for velocity (or at least redshift) as a function of distance, and much of the secondary literature assumes that this is what Lundmark found.

4. THE CALIBRATION OF THE CEPHEID AND RR LYTE DISTANCE SCALES

Both Hubble (Edwin) and Hubble (Key Project Team) ultimately tied their numbers to the absolute brightness of Cepheid variables as a function of pulsation period. To recapitulate, Henrietta Swan Leavitt first spotted a correlation of apparent brightness and period for variables in the Small Magellanic Cloud whose light curves showed the rapid rise and slower fall epitomized by that of Delta Cephei (Leavitt 1908). She made the relationship quantitative four years later (Pickering 1912, and you may well ask why she was publishing under the name of Pickering!). There were 26 stars with a dispersion of about half a magnitude around the mean line.

The next step was to assign an absolute magnitude to some one period. Ejnar Hertzsprung (1873–1967) took the first cut (Hertzsprung 1913), using 13 Milky Way Cepheids with measured proper motions and radial velocities and the method called statistical parallax. His zero point, the absolute magnitude of a hypothetical Cepheid with $P=1$ day, was $M_\odot = -0.6$. He used the word Cepheid generically (the first to do so), said the RR Lyrae stars were not part of the same correlation, placed the SMC at 10,000 pc on the basis of Leavitt’s stars, and totally ignored the possibility of interstellar absorption.

Next came Henry Norris Russell (1877–1917) and Harlow Shapley (1885–1972), who investigated the galactic distribution of eclipsing variables and Cepheids. They agreed that the RR Lyrae stars were not part of the same population as the disk Cepheids (Russell and Shapley 1914), and endorsed interstellar absorption at a level of about 2 mag per kpc (as then proposed by King 1914). Shapley later disowned the paper (see Fernie 1969, a detailed discussion of the history of the Cepheid period–luminosity relation).

Shapley’s first solo words on the subject (Shapley 1918, 1919) already have firmly established the (roughly) 1.5 mag calibration error that it took 30-some years to fix. Fernie has looked at the original data and finds that neglect of absorption, small-number statistics and poor data, and neglect of the contribution of galactic rotation to proper motions all enter. Unfortunately, as one tries to fix the middle problem by using larger, hence more distant, samples of stars, the first and third get worse. Shapley tied the disk Cepheids, globular-cluster Cepheids, and RR Lyrae stars into a single P-L relation. Because globular clusters are seen largely at high latitudes, they are not much absorbed, just nicely compensating for their Cepheids being fainter than disk ones. This of course illustrates Gingerich’s moral “consistent doesn’t mean correct.”

In yet another “foolish consistency,” Doig (1925, 1926) had used the apparent brightnesses and periods of the Cepheids S Normae in NGC 6087 and U Sgr in M25 as a check on the cluster distance he had determined by spectroscopic parallax. It was 30 years before most of the astronomical community recognized the existence of Cepheids in open star clusters or that the pairings might be useful. At one time, the supposed total absence of Cepheids from open clusters was regarded as a remarkable statistical anomaly!

5. EDWIN P. HUBBLE, 1925–1936 AND BEYOND

Hubble’s paper establishing the existence of external galaxies from the presence of Cepheids in them was read at the 1924–25 AAAS meeting by Russell (with ground work by Joel Stebbins) and won a modest prize (Berendzen et al. 1976), shared with Dayton Miller for the disproof of special relativity (Mermin 1996). Miller had done a Michelson–Morley experiment and got it wrong. One can image a committee made up of equal numbers of Young Turks and Old Fogeys coming to such a compromise on “best paper of the meeting!” Hubble (1925, 1926) published his extragalactic Cepheids rather slowly and cautiously, possibly because of the continuing discrepancy with van Maanen’s rotating spirals (see Trimble 1995 for details of this and other pre-expansion issues, which are also discussed at length in Smith 1982, Berendzen et al. 1976, and other secondary sources). In any case, the New York Times (in its pre-Walter-Sullivan days) had already scooped the technical journals, writing on 23 November 1924 (p. 6): “Dr. E. Hubble confirms view that spiral nebulae are stellar systems” (notice the title, initial, and correct plural, no longer part of the house style of most newspapers!). Another early appearance of
Hubble's Cepheids is often wrongly cited (e.g., by de Vaucouleurs 1982) as PASP, 5, 261, 1925. In fact, Vol. 5 of PASP appeared in 1893 or thereabouts. Probably the short-lived Publications of the American Astronomical Society is meant, but I have not checked this.

This brings us to 1929 January 17, when Hubble submitted "a relation between distance and radial velocity among extra-galactic nebulae" to the Proceedings of the National Academy of Sciences (Hubble 1929). His graph (Fig. 2) extends only to 1100 km s$^{-1}$ and does not include NGC 7626 at +3779 km s$^{-1}$ reported by Humason (1929) on the previous page of the same volume (though it has often been said that the more distant point "guided his eye" in drawing a line through the ones he shows). The vertical axis is labeled "velocity," but the units are given as km, not km s$^{-1}$! And the X axis extends all the way to 2 Mpc.

Why was Hubble's correlation believed when Wirtz, Lundmark, and so forth had not been? Partly it was the manner of the man (Sandage 1989). But, in addition, he had chosen Cepheid variables as his basic distance indicator. This meant, as we have already seen, that his distances were going to be wildly wrong. But they would be consistent. He had Cepheid light curves (e.g., Fig. 3) for the six closest galaxies (SMC, LMC, M31, NGC 6822, 598, and 5457) and used them to conclude that the brightest stars of any kind had absolute photographic magnitudes of $-6.3$. The next 13 galaxies were plotted using this assumption, and the last four, in the Virgo cluster, on the assumption that an average galaxy has $M_v = -15.2$.

Humason's ever-increasing velocities extended the relationship to larger and larger distances (Hubble and Humason 1931; Humason 1936; Hubble 1936), but did not change the slope by more than the traditionally allowed 1σ around Hubble's (1929) value of 500±60 km s$^{-1}$. The 1929 paper mentions 465, 513, and 530; the graph shows lines for 465 and 513. A modern eye examining the plotted points inevitably concludes that Hubble was perfectly honest about the random errors of the result. The problem lay, as nearly always, in systematic errors.

Much of what was written about Hubble from his own time down to the present (including things written by or directly quoting Hubble) should be believed only with caution. Christiansen (1995) has gone back to old newspapers, school and college records, and town, state, and national archives and reconstructed a more objective record. None of the changes (in France as World War I ended but probably not wounded; studied law but probably never practiced; on various athletic teams but not always the star) make the slightest difference to the astronomical significance of his work. The pictorial record confirms that in early and middle
life Hubble was tall, athletic, and good-looking in the style associated with the heroes of Horatio Alger books (complete with cleft chin, shared with his sister Lucy; it is usually carried by a single dominant gene). His impressive physique, however, provided no protection against a massive heart attack not long after the 200-inch telescope went into regular operation and death from a stroke before his 65th birthday. Hubble spent the latter part of World War II at Aberdeen Proving Ground in the ballistics program. Christiansen (1995) makes him sound pretty useless in the job, but this is not entirely the impression I received in talking with Dorrit Hoffleit, who worked under him there (one of the very few Ph.D. women in the group).

Hubble was apparently not attracted by the politics of scientific societies, in contrast to Shapley, who served terms as president of the American Astronomical Society, the American Association for the Advancement of Science, Sigma Xi, and probably other things. Hubble was elected to two terms as president of the Commission on Nebulae of the International Astronomical Union (later called nebulae and star clusters, then extragalactic nebulae, and now galaxies). But even this relationship was not entirely a happy one. At its 1925 meeting, the Commission turned down Hubble’s scheme for the classification of galaxies in favor of one with less conspicuous evolutionary implications. And, the Cleveland of his commission, he was the only President ever to serve two non-consecutive terms. The IAU now meets triennially and Commission Presidents are changed at the end of each General Assembly. In early years, the pattern was less regular. The Presidents and the numbers of the GAs that elected them (with years when needed) were Bigourdan (founder), Slipher (1, 2; 1922, 1925), Hubble (3, 1928), Shapley (4, 5, 6; 1932, 1935, 1938), Hubble (7, 1948), Baade (8, 9; 1952, 1955), Mayall (10, 1958), Lindblad (11), Minkowski (12), McVittie (13), E. M. Burbidge (14), Holmberg (15), Markarian (16), Westerlund (17), V. C. Rubin (18), van der Kruit (19), Tamman (20), Khachikian (21), and Trimble (22, 1994). Hubble (despite the disagreement over galaxy classification) was elected to the Commission in 1925 at Rome. Of the protagonists who appear later, Sandage and de Vaucouleurs were elected in 1955, van den Bergh in 1961, and Tammann in 1974 (and I was co-opted in 1982 to chair a revival of the supernova working group). It is, so far, the only IAU Commission ever to have had as many as three female presidents.

6. INTERMISSION

Sidney van den Bergh and I have made some effort to ascertain when Hubble’s name first became firmly attached to various items. The phrase “Hubble’s relationship” appears first (Tolman 1929b) and “Hubble’s law” soon after (Milne 1933). The number that Hubble (1929) himself had called $K$ appears first as “Hubble’s factor” (Haas 1938), then as “Hubbleschen Expansions-Konstante, $\alpha$” (Behr 1951) and the English equivalent “Hubble’s constant” in 1952 (Bondi 1952). Robertson (1955) appears to have been the first to use the symbol $H$.

A most important advance in the period between Hubble’s and Baade’s values of $H$ was the general recognition that interstellar absorption of light, at a level of about one magnitude per kiloparsec in the galactic plane, was a widespread and important phenomenon. Although absorption in individual dark clouds had been known since the time of Herschel or before and the possibility of general absorption discussed for decades, credit goes rightly to Robert J. Trumpler (1886–1956). He plotted apparent angular diameter versus apparent magnitude for a number of open star clusters (Trumpler 1930). The relationship was a straight line for nearby clusters, but soon curved downward, indicating that distances implied by apparent magnitudes were too large. The clusters looked faint partly because they were far away but also partly because their light was being absorbed.

Interstellar absorption was not coupled into the Cepheid period–luminosity relation nearly as quickly as you might expect. Several dozen astronomers, most of whose names are still familiar, took one step forward and two steps back in the next two decades (chronicled by Femie 1969). Among the forwards one might note Boris Gerasimovich (1888–1937), who thought that the difference between his disk-Cepheid distance scale and other people’s RR Lyrae-dominated scales might be a result of absorption (Gerasimovich 1934), but then did nothing about it (history did not allow him much time). Boris Kukarkin (1949) firmly decoupled the classical Cepheids and RR Lyraes, but also without exploring the larger-scale consequences. Best directed of all was Meunier (cited with different years, but the copy I have seen says the paper was submitted in March 1945 and published in 1946). He started from scratch, admitting the probability of interstellar absorption, concluded that it was about 1 mag per kpc in the galactic plane, and said that the Cepheid zero point should be moved 1.0–1.1 mag brighter. He also did not explore the cosmological consequences of a Cepheid recalibration, but his Hubble constant would have been about 325 km s$^{-1}$.

Lundmark appears on our stage one last time, saying firmly that there is something wrong with M31, because its distance based on novae and globular clusters, if their brightnesses were the same as Milky Way examples, was a factor of 2 larger than the distance implied by Shapley’s Cepheid calibration (Lundmark 1946, 1948, 1950; this last reporting event at the 1948 Zurich IAU).

Outside the Local Group of galaxies, was lurking another, independent factor of two, first recognized by Behr (1951). He had discovered that assuming a fixed absolute magnitude for the brightest galaxy in a cluster would inevitably lead to apparent distances smaller than the real ones by increasing amounts as you look farther away. This is not a reinvention of Malmquist bias (under which assuming a constant value for average galaxy brightness leads you astray). Rather, it is a heuristic pre-discovery of what we now call the Scott effect, as is made clear in the pioneering discussion by Elizabeth L. Scott (1956, an extended meeting abstract). The point is that you need to be able to see a certain number of galaxies to recognize a cluster. As distances increase, clusters have to be richer and richer to have enough bright galaxies to be recognized. And, eventually, you get to clusters richer than any in your local sample, so that they contain rare, super-bright galaxies not present in the calibrating data.

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System
Behr's discussion implies a Hubble constant near 240 km s\(^{-1}\), without any Cepheid recalibration.

Within the same time frame, an older "long-short" disagreement was being resolved. It pertained to time scales, rather than distances, and was largely the fault of Sir James Jeans (1929, e.g.). From the time of Archbishop Ussher through the 19th century, the world had been aging rapidly. Jeans carried this to extremes with considerations of stellar dynamics. His time scales came from attempting to relax star clusters and elliptical galaxies with only close two-body encounters and from turning initially circular binary star orbits into eccentric ones (quite the opposite of our modern concept of binary orbit evolution) also with close encounters. He also assumed the old Russell "giant and dwarf" theory of stellar evolution, in which stars progress down the main sequence from early to late types, losing mass as they go via annihilation (presumably of protons with electrons). His time scales were all near 5\(\times\)10\(^{12}\) yr, and he saw no objection to all nebulae being that age, and the whole universe having been created in an instant 10\(^{-12}\)-13 yr ago.

Meanwhile, radioactivity in Earth rocks and alternative scenarios of stellar evolution were yielding numbers in the range 10\(^{8-10}\) yr (see notably Eddington 1926). Cosmic expansion with a large Hubble constant naturally weighed in on the short time scale side of the balance. At a 1935 meeting on the subject reported in PASP, Vol. 47, Tolman (p. 202) makes this point. Robley and Evans (p. 199) and Beno Gutenberg (of the discontinuity, p. 200) had looked at Earth rocks, meteorites, and the lunar orbit, concluding that the solar system cannot be more than about 5 Gyr old. And Gerard Kuiper (p. 201) explained what he thought Jeans had done wrong. Jeans was not immediately persuaded and continued to advocate the "long" time scale in later books, but changed his mind shortly before his death in 1946 (Hufbauer 1994).

As early as 1933, Eddington had expressed doubts about the Milky Way being truly much larger than all the other spiral nebulae, as had to be the case with the distance scales of 1925–1951. He was about to be vindicated, though no longer in a position to say "I told you so."

7. ROME, 1952

Contracts for building the 200-inch (5-m, Hale) telescope at Mt. Palomar in 1928, and the first useful photons were gathered about 20 yr later (after a worrisome shakedown period of somewhat uncertain length). All experts had predicted that the 200-inch should comfortably image RR Lyrae variables in M31. Shapley's calibration put M31 at a distance modulus of 22.4 mag, corresponding to a distance of 275–300 kpc, with some allowance for light absorption. The RR Lyraes, thought then to have \(M_{pg}\) near 0, should have appeared at \(m_{pg}\) of 22.4 or thereabouts. They did not. As Walter Baade (1893–1960) reported at the IAU General Assembly in Rome, only the very brightest Population II stars—the tip of the red giant branch—were visible on limiting exposures.

Within Milky Way globular clusters, the RR Lyraes are 1.5 mag fainter than the red giant tip. Thus Shapley's calibration had to be wrong by the same amount. Baade described the implication not as a change in Hubble constant, but as an increase in the age of the universe from 1.8 to 3.6\(\times\)10\(^7\) yr.

Confirmation was already at hand. Andrew David Thackeray, working at the Radcliffe Observatory in South Africa, had just discovered, and reported at the same Rome IAU, three RR Lyrae stars in the Small Magellanic Cloud. They too were 1.5 mag fainter than predicted by the Shapley calibration. RR Lyraes in the Large Magellanic Cloud followed the next year (Thackeray and Andriaan Jan Wesselink 1953). RR Lyrae variables in the SMC had actually been reported much earlier by Shapley, but every period he gave was wrong (too short by factors 2–4), and none of the stars are actually RR Lyraes (Smith 1982, p. 123). The transactions of the Rome IAU were edited by Oosterhoff (1954), and Baade told his story in a Bruce Medal Address (Baade 1956).

Once it was safe to admit in public that Shapley might have been wrong, recalibrations of the Cepheid distance scale multiplied in a style owing something to both rabbits and sheep. Fermie (1969) reports 19 such efforts between 1953 and 1959, and the list is probably not complete. Curiously, one of the recalibrations comes from Shapley himself (Shapley and McKibben Nail 1954). Gerard Henri de Vaucouleurs (1918–1995) first enters extragalactic space in this period, placing the Magellanic Clouds at an average distance of 52 kpc (de Vaucouleurs 1955).

8. FURTHER DOWNWARD REVISIONS

Work on recalibration of Cepheids has continued to the present time (including, of course, the HST key project), but the next major changes in Hubble constant came from other considerations and are largely associated with the name of Allan Sandage (see Fig. 1). Sandage received his Ph.D. from California Institute of Technology the same year, 1953, that Hubble, whose assistant he had been, died. We all know Sandage's middle initial is R. (for Rex) and are forever sticking it in when citing his papers. He himself hardly ever uses it, undoubtedly feeling that there are not an enormous number of other A. Sandages from whom he needs to be distinguished. If you doubt the correctness of this, check any large phone directory; Sandage, de Vaucouleurs, Tamman, van den Bergh, and Slipher are completely unknown to Orange County, but there are two Hubbies and two inches of Trimbles.

Humason, Nicholas Mayall, and Sandage (1956) provide the next landmark. Their massive paper deals primarily with the accurate determination of nebular magnitudes and redshifts. But, tucked away in an appendix, is an estimate of the Hubble constant. It rests on several milestones significantly different from those adopted by Hubble. The brightest single stars are placed at \(M_{pg} = -8.5\) (based on examples in M31 and M33, many of them Hubble–Sandage variables) and brightest galaxies in clusters are set equal to M31 at \(M_{pg} = -19.92\). This moved the Virgo cluster out to a distance modulus of \((m - M) = 29\), which, in combination with an average velocity for the cluster of 1136 km s\(^{-1}\), yields a Hubble constant of 180 km s\(^{-1}\) Mpc\(^{-1}\). They believed this to
be correct to within 20%, assuming as usual that the main errors were statistical ones of 0.2–0.3.

Two years later, the rules changed again. Just which astronomical objects are point sources and which are fuzzy is an ancient topic of dispute (most recently for quasistellar objects observed with HST). Hubble had been wrong about some of these. Sandage (1958) pointed out that some of his "bright stars" in galaxies of Virgo and beyond were really entire H II regions and their illuminating central stars and clusters. Thus they were much brighter than previously supposed. This moved their host galaxies outward another factor of 2–3. Sandage proposed \( H = 75 \text{ km s}^{-1} \text{ Mpc}^{-1} \), to within a factor of 2; I believe this may have been the last completely non-controversial—and honest—value published, owing to rapid shrinkage of error bars on all sides.

A number of astronomers expressed preferences for values between 100 and 200 km s\(^{-1}\) Mpc\(^{-1}\) during this period (sometimes on the basis of no new data at all). These include McVittie (1959) advocating 143–227, Sersic (1960) recommending 125±5, Holmberg (1958) favoring 134±6, and Ambartsumyan, speaking \textit{ex cathedra} at an IAU invited discourse (1961) in favor of the range 70–100 km s\(^{-1}\) Mpc\(^{-1}\).

Van den Bergh's (1960a,b) first two published calibrations were 100±20 and 120±25. De Vaucouleurs (1964) first appears on the Hubble constant stage drawing attention to an asymmetry associated with the Virgo cluster, such that \( H = 100 \) in its direction and 125 elsewhere. This corresponds to a Virgocentric infall (as he did not call it) of 250–300 km s\(^{-1}\) Mpc\(^{-1}\).

9. THE BATTLE LINES HARDEN

IAU Symposium 15 in 1961 (in connection with the Berkeley General Assembly) focused on Problems of Extragalactic Research (McVittie 1962). Sandage (1962) attempted to arrive at a consensus value by averaging his own 75 and 82±18 (from H II regions) with the numbers mentioned just above as found by van den Bergh, Sersic, and Holmberg. Unfortunately, he did not use the numbers they had published, but his own renormalization of them to his Local Group Cepheid distance scale, which removed about half of the real spread (or discrepancy, depending on your point of view). This was presumably not in accord with the views of the people thereby renormalized, but resulted in the choice \( H = 100 \) km s\(^{-1}\) Mpc\(^{-1}\) being very widespread for a decade or two. The convenient hiding of remaining uncertainty in the parameter \( h = H/100 \) arises from that brief era of good feeling. Sersic (1962) was already aware that a Hubble constant near or a bit in excess of 100 km s\(^{-1}\) Mpc\(^{-1}\) would get you into age trouble with the globular clusters.

By the time of Vol. 10 of the Kuiper compendium (Sandage et al. 1975), it was necessary to have two separate reviews of the cosmic distance scale. Van den Bergh (1975, but actually writing in September 1969) settled on 95±15, while Sandage (1975, written in August 1972) declared for 55±5, and he has since at times come down as low as 42±11 (Sandage 1988).

As we enter the modern era, the Sandage and Tammann series of "steps toward the Hubble constant" converged 20 years ago (Sandage and Tammann 1976) but was followed by many more papers addressing various aspects of the distance scale and age problems, always finding numbers near 50–55 km s\(^{-1}\) Mpc\(^{-1}\). De Vaucouleurs (1979) ended his next assault at 100±10. He took up cudgels and Sosies again almost immediately and laid them down only upon command of higher authority next to a Hubble constant of 87.3±1.1 (de Vaucouleurs 1993). Sadly, the last word I heard from him on the subject was an angry one, in response to an honest, but frivolously phrased, remark in a review article (Trimble and Leonard 1994). Van de Bergh has published an interestingly wide range of Hubble constants in recent years, but opined that "on balance, available evidence suggests that \( H_0 \approx 75 \) km/sec/Mpc" (van den Bergh 1994).

Is there a Trimble value of the Hubble constant? Well, yes, sort of. But it is maximally wishy-washy. I always begin introductory classes and public talks on cosmology and the early universe by saying "between 10 and 20 billion years ago," and I calculate everything for more advanced classes with all the unhappiness hidden in various powers (with luck usually the right ones) of \( h \). I think the correct value probably lies in the range 50±30 km s\(^{-1}\) Mpc\(^{-1}\), but confess to having used permanent ink to set my debate Hubble-meter at 33.

Note added in proof: Albert E. Whitford's first name and initial were given incorrectly in my paper from the previous Curtis–Shapley restaging (PASP, 107, 1133). I apologize to him and to any reader who might thereby have been confused!

REFERENCES

Ambartsumyan, V. A. 1961, Trans. IAU, XIB, 150
Behr, A. 1951, Astron. Nach., 279, 97
Brush, S. 1993, Perspectives Sci., 1, 565
de Vaucouleurs, G. 1982, The Cosmic Distance Scale and the Hubble Constant (Canberra, MSSO and ANU)
Djorgovski, S. G. 1995, private communication
Doig, P. C. 1925, JBAA, 35, 202
Doig, P. C. 1926, JBAA, 36, 60 and 202