Title
TIME SCALES FOR ACHIEVING ASTRONOMICAL CONSENSUS

Journal
International Journal of Modern Physics D, 17(06)

ISSN
0218-2718

Author
TRIMBLE, VIRGINIA

Publication Date
2008-06-01

DOI
10.1142/s0218271808012590

International Journal of Modern Physics D
Vol. 17, No. 6 (2008) 831–856

World Scientific
www.worldscientific.com

TIME SCALES FOR ACHIEVING ASTRONOMICAL CONSENSUS

VIRGINIA TRIMBLE

Department of Physics and Astronomy,
University of California,
Irvine CA 92697-4575, USA

Las Cumbres Global Telescope Network,
Goleta, California, USA

Received 7 April 2008
Communicated by IJMPD Editorial Office

The history of science can be recounted in many ways: by addressing the work of one person or school; by starting with the ancients and working chronologically up to the present; by focusing on a particular century; or by tracing a particular important idea as far back and forward as it can be found. The present discussion does none of these. Rather, it adopts the ordering of a standard introductory astronomy textbook, from the solar system via stars and galaxies, to the universe as a whole, and in each regime picks out a few issues that were controversial or wrongly decided for a long time. For each, I attempt to identify a duration of the period of uncertainty or error and some of the causes of the confusion. This is surely not an original idea, though I am not aware of having encountered it elsewhere, and it is not one that is likely to appeal to most 21st century historians of science, for whom the question "Who first got it right?" is not necessarily an important, or even appropriate, one. Some of the stories have been told as historical introductions to conferences and are here summarized and brought up to date. Others I had not previously addressed.

Keywords: Astronomy; history; solar system; stellar evolution; Milky Way; cosmology.

1. Introduction

To those of us who remember them, the 25 years from the announcement of the existence of gamma ray bursts[1] to their firm identification with very powerful events occurring only very rarely in any one galaxy (and, therefore, typically to be seen only at large distances)[2,3] seemed like a very long time, especially to the subset who also remember how quickly pulsars had been associated with rapidly rotating, strongly magnetic neutron stars[5–9] only a few years before. Indeed, as is not obvious with the format chosen by this journal for citing references, the first three of these actually have the status of predictions, in the sense that the correct explanation was in the theoretical inventory before the corresponding objects were found.

The rapid proliferation of models for GRBs was also not unique — even in the pulsar case there had been time for brief consideration of white dwarfs as well as neutron stars as the underlying bodies, and of pulsation and orbits as alternatives to rotation as the time mechanism, in connection with the latter of which a gifted theorist, relatively new to astronomy, succeeded in rediscovering the Roche limit.[10] The sheer number of ideas put forward for GRBs was, however, probably unprecedented, and even the list of 118 proposed explanations up to 1994[11] was not quite complete. More, of course, followed in the next few years. Distressing in retrospect, though also not uncommon, was the extent to which the astronomical community converged on a fundamentally wrong explanation, involving surface events on old neutron stars in the plane or halo of the Milky Way.

I have told aspects of the GRB story before, most recently as the historical introduction to a symposium celebrating the first couple of years of burst data from the SWIFT satellite,[12] and will focus here on some other stories from the history of astronomy. When theory has led to observations, we generally speak of prediction; the converse is called discovery and explanation. The division is not a clean one, and you might at least want to distinguish cases where prediction has motivated observation and ones where the discovery was made without knowledge of earlier predictions. The cosmic microwave background is a classic of this latter sort.

Let us begin with a few definitions. "Discovery" means the moment when at least a few people agree that something has been observed that requires an explanation. Thus supernovas were discovered by Baade and Zwicky in 1934,[13] not by a hypothetical Zinjanthropan named Zog, who saw a galactic one in the year 2,345,678 BP. A problem counts as "solved" when most of the community has converged on an explanation and the convergent point has held down to the present. Notice, looking back over your shoulder, that these definitions would say that the correct value of the Hubble constant became a problem in 1929, though Hubble himself called it K (H becoming ubiquitous only after his early death), and that the problem was solved for epochs of varying lengths by Hubble himself,[14] by Walter Baade ($H = 250\,\mathrm{km/s/Mpc}$),[15] by Humason, Mayall, and Sandage ($H = 180$),[16] and by Sandage ($H = 100$),[17] though use of $H = 100\,\mathrm{km/s/Mpc}$ persisted as a computational convenience many years beyond the time when Sandage[18] began saying that 50 or 55 was closer to the truth. The present convergence around a value close to 70 might also be regarded as subject to revision, though it is clearly the winner at present, appearing in the most-cited astronomical paper of 2001[18] (Sec. 5.2).

Notice that both discovery and solution involve something very much like voting by scientists collectively until there comes a time when the accepted explanation has passed a very large number of observational (or experimental) tests and failed none. Consensus also matters when theory leads, often with a good many false alarms along the way, to an accepted discovery. Parallax (predicted by various

Greeks and found by Bessel, Struve, and Henderson in the 1830s) is an extreme example.[23] Extra-solar-system planets were also widely expected and widely (but wrongly) reported before observations in the early 1990s counted as discoveries, largely because they could be, and were, confirmed.[20–22] Of course, the confirmation must also hold down to the present to count; a remarkable number of folks reported seeing Vulcan in the second half of the 19th century.[24]

There is no way to weave the following items into a single, coherent story, and the ordering is simply that of a standard introductory astronomy book, near to far, with a bit of cross-referencing.

2. Some Spotty Stories about the Solar System

2.1. *The ears of Saturn*

Galileo, fresh from the triumph of finding his Jovian moons, turned to Saturn and saw protuberances on either side, which he called "ansae," or ears, and initially supposed might also be moons. But they disappeared in 1612, reappeared in 1613, and could not be fit by any periodic motion that he could see. We then skip ahead a generation to Christian Huygens, who was 13 when Galileo died and who started observing Saturn with a Keplerian telescope of his own design and construction in spring 1655. By November, the ansae were invisible again, reappearing through summer and fall 1656. He published a brief tract explaining them as a ring inclined to the plane of the ecliptic by about as much as our equator, so that disappearance happened twice per Saturnian synodic period as we crossed the ring plane. Huygens, who was a Cartesian, concluded that the ring would be stable in the Saturnian vortex.[25,26] Note that Ref. 25 is a two-volume encyclopedia, where the relevant information is to be found under the name of the person mentioned in the sentence; and Ref. 26 is a comprehensive history of astronomy and some iterating between text and index may be necessary to find some of the items. The ring appears at greater length in Huygens's 1659 monograph *Systema Saturnium*.

Skip ahead again 200 years to 1857, when James Clerk Maxwell (in a prize essay written at Cambridge) showed that the rings (by then multiple) must be made of small pieces, since any solid or fluid ring system would be torn apart by Saturnian tidal forces. We touch down one last time to watch James Keeler at the Allegheny Observatory in 1895 recording a spectrum of the rings with evidence for differential rotation, showing that indeed the particles further out moved more slowly, while a solid ring would have the most rapid motion furthest out.

I count this as an incompletely told story because of not having come across any evidence for how much, if at all, astronomers between 1613 and 1657 worried about the nature of the ansae, or those of the next two centuries about the physics of the rings. Establishing the composition of those small particles (water is important) and the dynamical evolution of the present system (lifetime much less than the age of the solar system) counts as work in progress.

2.2. *The Great Red Spot on Jupiter*

That the nature of the Red Spot was long puzzling is revealed by an apocryphal mid-20th-century tale of Newton describing universal gravitation to a learned society (this never happened) only to have a fussy elderly member ask, "Well, young man, and can your theory explain the Great Red Spot on Jupiter, one of the most puzzling phenomena of the solar system?"[27]

The physical explanation appears to belong to the class of "sporadic convergences around transient ideas," since the spot was: (a) something floating in a fluid medium to Russell, Dugan, and Stewart (Ref. 28, a classic early 20th century text, and a convenient all-purpose reference to mainstream opinion in its day); (b) a Taylor column when I was a graduate student (halfway from Russell *et al.* to the present); and (c) a poorly characterized anticyclonic feature in recent times. The spot now is not nearly so red nor so conspicuous as it was in the late 19th century. Popular recognition of this has been obscured by the greatly enhanced colors of Voyager images released by NASA in the 1970s.

But the historical puzzle is who discovered the thing. Cases can be made for Cassini in the 1600s, if the Permanent Spot was an early appearance of the GRS, for someone in the early 1700s if the red color of a Jovian spot in a 1711 painting was accurate, and for Carr Pritchett in 1878, plus many others shortly before and after, as part of a decade in which the surface features of Jupiter had apparently been changing very rapidly, and photography was not quite ready to adjudicate among drawings. Convergence happened at least with respect to observed properties within a year or two — nearly a perfect ellipse, bright color (rosy to dirty flesh, depending on the observer), and rapid proper motion relative to other Jovian atmospheric features in longitude but stable in latitude south of the equator.

The first Jovian photograph, by Andrew Common, confirmed the size, shape, and location though the color, of course, was gray. Since I have taken most of the story from Hockey's monograph with "Before Photography" as part of its subtitle,[29] the story naturally ends here.

Duration: a bit more than 90 years from firm recognition of the GRS to its being subsumed into the more general problem of atmospheric circulation patterns on Jupiter in the wake of Voyager images.

2.3. *The advance of the perihelion of Mercury*

A relatively well-known story, because it ends with the triumph of general relativity, this one also involves some apparent spots seen on the face of the Sun. The year after his 1846 prediction of an extra-Uranian planet (found almost immediately by Galle in Berlin), U.-J. J. Le Verrier[25] resolved to chase down every discrepancy he could find between predicted and observed positions in the solar system. By 1859, the outstanding one was a rotation of the orbit of Mercury at a rate 38″/century faster than could be explained by Venusian and other known perturbations.[30] (The modern value is 43″/century.) Le Verrier wondered about this loudly enough to be heard in

Orgeres, 75 miles SW of Paris, where physician Edmond Lescarbaulty had recorded a dark spot moving across the face of the Sun faster than solar rotation could account for, in the spring of the same year. Le Verrier found the observation credible and coined the name Vulcan, compiling a list of apparent observations of it both before and after 1859.

The Vulcanic explanation was at least a majority view for a couple of decades, though non-detection during solar eclipses and the impossibility of fitting all the reports with a single orbit led to gradual doubt. By 1900, Simon Newcomb[31] attached at least equal probability to some deviation from Newtonian gravity, though there was worry that it ought also to show up in the orbit of Venus (it does, but at a much lower level of $8''$/century) and of the Moon around the Earth, There is also a relativistic effect expected for the Earth–Moon system, but it is much smaller and currently lost deep in other uncertainties about the theory of lunar motion. If you get out a general relativity text, you can check my arithmetic, which led to something like $0.06''$/century.

Yet another possibility was a nonzero quadrupole moment for the Sun, which was to cause trouble later. Einstein's 1915 paper introducing GR[32] needs no introduction to any reader of these pages, and one might, therefore, claim a duration of 1856–1915 for the Mercurial conundrum. If, however, you ask when the full mainstream community converged on the GR answer, the dates become much fuzzier, Russell *et al.*[28] were confirmed relativists by 1926; but Heber Doust Curtis (who was right in the Curtis–Shapley debate about the existence of external galaxies) had, to the end of his life in 1942, "not much use for that fellow Einstein," according to the last of his students.[33]

And soon after, by the stretched-out time line of this discussion, Robert Dicke was attributing part of the perihelion advance to a solar quadrupole, which he thought he had detected, thereby favoring a different, scalar-tensor theory of gravity that he had put forward in 1961.[34] That particular threat has faded with downward revision of the solar quadrupole. (Dicke's measurements had been sensitive to faculae near the equator.) But alternative theories of gravity have proliferated in recent years, partly as ways of incorporating dark energy and partly as attempts to provide a renormalizable theory of gravitation that might some day be unified with the other three forces. No consensus so far; but any such theory will be required to reproduce the solar system results.

3. Some Stellar Stories

3.1. *How far are the stars?*

Introductory astronomy books all know the "what and where and when, and how and why and who" answers to this. The what is parallax, the when 1835–39, how = precision astrometry, and why = final demonstration of the "Copernican hypothesis," though the where and who get somewhat divided among Bessel/Konigsburg (61 Cygni), Struve/Dorpat (Vega) and, among the cognoscenti, Henderson/South

Africa (Alpha Centauri).[23,25,26] The three measurers all agreed on distances in excess of 1 parsec, with the implication that most stars would be a good deal more distant. They thereby ended a search that had been (very sporadically!) in progress since the beginnings of Greek natural philosophy, though by the time parallax was actually measured no one was surprised by either the large distances or the verification of a Sun-centered solar system.

The ancients nearly all agreed that the stars were more distant than any of the planets (though Hildegard of Bingham mixed them among hail, lightning, the Moon, and inner planets — picture in Ref. 35), typically not much more distant, though it is sometimes difficult to tell the extent to which this was driven by finite paper or parchment size.

Pre-telescope observers also agreed that a rather faint star had an angular diameter near 1 arcmin.[23] Thus a "stars are suns" argument would put them at tens of astronomical units. This would seem to be what Kepler had in mind when he nested his platonic solids to set the positions of the planets and put the stars not far beyond the last planetary sphere; and it was presumably the consensus view for most of recorded history, the absence of parallax being taken to demonstrate the correctness of Ptolemaic cosmology.

There was, however, already a contradiction built in. Tycho's upper limit of $0.5'$ heliocentric parallax put the stars at least 700 times further away than Saturn, driving him to a compromise system in which the planets orbited the Sun, but the Sun circled the Earth. The stars could then still be rather close to us and not enormous. Even the first telescopes, which Tycho (1546–1601) did not live to see, but Kepler (1571–1630) did, cut stellar angular diameters down to a few arcseconds, temporarily solving the "bloated star" problem. Distances equivalent to thousands of Earth–Sun distances then made sense for another century or so.

But we were still more than a century away from actual parallax measurements when two further considerations agreed in pushing the stars out to millions of AU. First, James Bradley[23,25] in 1729 set an upper limit near 1 arcsec to the parallax of Gamma Draconis, simultaneously demonstrating aberration of starlight and, thereby, the orbital motion of the Earth around the Sun. Second, a succession of savants (beginning with Newton [d. 1727]) published various comparisons of the apparent brightness of the Sun by day and the stars by night. John Michell in 1784,[25] using Saturn as an intermediary, was particularly successful. Using this method led William Herschel a bit later to construct his Sun-centered Milky Way with the closest stars already more than 200,000 AU away and the fainter, more distant ones a thousand times or more farther. His statement that he had seen light that had been traveling from stars for millions of years, though well-documented, I still find mysterious. A sun at 500,000 pc has an apparent magnitude of +23.

In any case, the community had already enjoyed a hundred years of consensus on large stellar distances by the time Bessel, Struve, and Henderson announced their results, and I am not quite sure they would have received Nobel Prizes if the concept had existed then. Struve and Bessel were awarded Gold Medals by the Royal Astronomical Society (Struve before parallax, Bessel after), Henderson not even that. Other winners from the same period include names you will recognize (John Herschel) and others you may not (John Wrottesley), though you can probably guess what Lt.-Col. George Everest discovered.

Duration of the problem: more than 2000 years in the form of the search for parallax perhaps a century of agreement on far too small a scale for either Ptolemaic or Copernican structure.

3.2. *The source of stellar energy*

This was declared by Newcomb[31] among others to be the single most important unsolved problem in astronomy a century ago. It is another fairly-well-known story, though one with a curious detour in which much of the community worried about the consensus answer 30 years after it had been accepted in 1939. The story is not yet fully told, and I note here two items cherry-picked since I told it last.[36]

The problem was, in a sense, born backward. James Hutton (1726–1796) and Charles Darwin (1809–1882) were asking for hundreds of millions of years for geological and biological processes well before Lord Kelvin (William Thompson, 1824–1907) began telling them in 1862 that they could have at best 20–100 million years[37,38] for a Sun powered by gravitational contraction. It is worth noting that Kelvin had almost simultaneously determined an age for the Earth, assuming it to have cooled from an initially high temperature without further energy input. Getting the same answer from two such different considerations gave him great confidence in it (somehow maintained to the end of his life) — and we should watch out for this sort of excessive confidence in current astronomy! Kelvin was also unreconciled to Darwinian evolution and went to his grave denying the possibility of transmutation of the elements and proudest among all his work of his calculations pertaining to the age of the Earth.

Kelvin's solar energy source was gravitational contraction, and, by way of additional chaos, he had scavenged the idea from Hermann von Helmholtz (1821–1894), and both had been anticipated by a decade or more by Julius R. Mayer (1814–1878, moderately well known to the history-of-chemistry community) and by John J. Waterston (1811–1883), not particularly well known to anybody, but memorialized by J. S. Haldane, both of whom had their papers rejected.[25,38] Nevertheless, gravitational energy was the only game in town for about half a century, during most of which it was seen as inadequate, hence Newcomb's[31] remark.

You all know that the answer is nuclear fusion, beginning with $4\mathrm{H} \rightarrow \mathrm{He}$, plus about 0.8% of the rest mass energy liberated in photons, neutrinos, and kinetic energy, and that Hans Bethe, at least, of the many folks involved received a Nobel Prize for it (1967). It seems to me, exercising extreme hindsight, that there were five definite steps that had to be taken on the way, though a couple of the time frames overlap:

(1) $E = mc^2$ (Einstein in 1905). That is, if some mass disappears, energy must pop out.
(2) One helium is less massive than four hydrogens, for which astronomers nearly always credit F. W. Aston (Nobel Prize 1922) and his Cantabridgian mass spectrograph. Indeed, his numbers were much better than previous ones, and helium is very difficult to tackle by chemical methods, but that 16 hydrogens was more massive than one oxygen had long been known. Venable[39] tabulates the weight of hydrogen as 1.0024 on a scale where atmospheric oxygen = 16, with credit to Frank Wigglesworth and to Meyer and Seubert (Lothar and Carl, if you should ever need to introduce them[40]).
(3) The stars (and Sun) are made mostly of hydrogen and helium,[41,42] a result accepted only gradually by the astronomical community. Eddington in 1926[43] thought that 7% hydrogen was the most you could tolerate (and he was a firm supporter of subatomic energy), and Russell still later held on to less than half by mass, though more by number.
(4) Barrier penetration, which allows two protons, or a proton + a C^{12} nucleus, to approach each other within the range of the nuclear force (Gamow, 1928, and Condon and Gurney the same year for alpha-particle emission, and Atkinson and Houtermans,[44] 1929, in a particle capture context).
(5) The detailed reaction sequences. We have all agreed since 1939 that there are two basic ones — the proton–proton chain and the CNO cycle — and most of the credit generally goes to Hans Bethe.[45] But again there are precursors. The idea of starting by trying to force two protons together reached Bethe from C. F. von Weizsacker. And using some heavier element as a catalyst to bring together four protons appears. Atkinson and Houtermans,[44] writing before the 1932 discovery of the neutron, also had to have their proton bucket scoop up two electrons along the way.

Two loose ends, however, remained, on the basis of which you might want to claim that full convergence did not occur for another 15 or even 60 years, beyond the 1939 date for reasonable understanding of hydrogen fusion in stars (and the outbreak of WWII, which delayed progress across the whole frontier of science!).

First, Bethe's nuclear physics was better than the solar model available at the time, which pegged the central temperature at 20 million K. At that temperature, the CNO cycle would provide virtually all the observed solar luminosity, because its reaction rate is a steeper function of temperature than is the rate for the p–p chain, as a result of the higher Coulomb barrier. Thus Bethe sensibly concluded that only stars much fainter than the sun would live on p–p fusion. But the temperature was too high (in turn partly because the composition was wrong — only hydrogen gives you two particles to exert pressure for every dalton of molecular weight). Curiously, the folklore does not firmly credit anyone for showing that the sun actually runs mostly on the p–p chain, though it was firmly established by the 1958 date of Schwarzschild's classic text.[46]

And then came the celebrated solar neutrino problem, A very-carefully-checked experiment, for which Ray Davis won a 2002 Nobel Prize in physics, recorded only about 1/3 as many high energy neutrinos as were predicted by the best models dating from the 1960s down to the present. John Bahcall, who was there from the beginning, has told the story in accurate detail[47] down to 1989, including a suggestion from Stephen Hawking that the correct source of solar luminosity might be accretion on a small central black hole rather than Bethe's nuclear reactions. No, it doesn't work any better than it did when Lev Landau put forward a similar idea in the 1930s.

But doubts continued to be cast for another 10 or 15 years until convergence finally occurred again, around the idea of matter-catalyzed neutrino oscillations (MSW effect), so that only about 1/3 of the high energy neutrinos produced by the sun reach us as the electron ones to which Davis was sensitive. This has now been well established by other neutrino detectors, sensitive to other energies. Minds shifted from blaming nonstandard astrophysics to non-standard weak interaction physics rather slowly, Bahcall's before 1989 and many of the rest of us around 1990 after reading Bahcall and Bethe's discussion,[48] though this was actually a bit before the most telling experimental data appeared.

The problem of the source of solar and stellar energy thus had a total lifetime of 130 years or more, with several intermediate stopping-points, but no convergence on a wrong answer once gravitational contraction was generally agreed to be insufficient.

3.3. *Coronium, nebulium, and other mythical elements*

Norman Lockyer's announcement of the discovery of a new element, helium, in 1868 as the source of a yellow line in the flash (chromospheric) spectrum of the sun was such a success that other observers were enticed to invent other -iums to account for other spectral lines not found in laboratory studies of known elements. Two came in 1869, when Thomas Young coined the name "coronium" for a 5303 Å line in the solar coronal spectrum and William Huggins picked "nebulium" for bright emission lines at 5007 and 4959 Å, which he had briefly thought might be signatures of nitrogen in Orion and other nebulae.[25,28] Casseiopeium, asterium,[40] and aldebarium also appeared at some point in print and were occasionally inserted between hydrogen and lithium in early periodic tables. There were also the metaelements decipium, phillipium, and mosandrum of William Crookes,[49] which even he eventually accepted as known rare earths and mixes of them for which correlation of spectral features and chemical properties is particularly difficult.

The problem, of course, was that there was really no place in the post-1900 periodic table for these astronomical elements, even less so after Moseley focused attention on the importance of atomic number versus atomic weight.[40]

Nebulium yielded up its secrets first, to the physical insight of Ira S. Bowen,[25] who considered known levels of doubly ionized oxygen and showed in 1927 that the green lines would be emitted by transitions not normally connected by radiation, because the photon had to carry off a possible but not probable quantum of angular momentum. Thus only from very dilute gas in a very large volume would you see the lines, called forbidden. A fairly clean 58 years, though nebulium (as opposed to some unknown element in an unusual state) was probably the best-buy hypothesis for only about the first half of the period. Nevertheless, Nicolson in 1917[50] and Arrhenius in 1922[51] were still using it in their scenarios for star formation.

Coronium held out for another 15 years, yielding also to deeper understanding of atomic structure, this time by Grotrian and Edlen,[25] who considered isoelectronic sequences and the wavelengths that might be radiated by heavy elements deprived of many of their electrons. Because there was a war going on at the time of their insights, word that coronium was, among other things, iron deprived of 13 electrons reached across the Atlantic in a paper by someone else entirely.[52] That rare conditions rather than a rare element must be involved had been obvious to Russell *et al.*[28] in 1926. The high ionization state implies that the unusual condition is very high temperature, and an assortment of heating mechanisms have been put forward from 1946 to the present. I think there is still some sorting out to do here among acoustic processes, microflares, MHD waves, and perhaps other physics; so, 73 years for the basic identification and another 66+ for full physical understanding.[113]

3.4. *Variable stars, non-traumatic*

Fadings of Algol were first reported in 1667 by Geminiano Montanari,[25] who, however, did not observe the star regularly enough to discover its periodicity. This was left for Charles Goodricke and Edward Pigott in 1781,[26,53] though Ismael Boulliau had announced an 11-month period for Mira back in that epochal year of 1667.[26] Boulliau suggested dark spots on a rotating star, an obvious extension of what a sufficiently sensitive observer of the sun might see from far away. Goodricke and Pigott began with eclipses by a dark planet, half as large as the star, and found this satisfactory for Beta Lyrae, but backed off when they realized that their asymmetric light curve for Delta Cephei could not easily be explained by eclipses. Pigott, after Goodricke's (1764–1786) very premature death, adopted a rotating, spotted star model for R. Scuti and then seems to have lost interest in astronomy in the last decade or two of his much longer (1753–1825) life.[25]

And there the matter rested for more than a century, though the eclipsing model seems to have been nearly forgotten in favor of rotation and spots (which could, of course, explain any light curve if the spots were allowed to move and vary like sunspots do). Eclipse models were revived by Edward Pickering in the late 19th century,[28] when he had spectroscopic evidence for stars orbiting each other more or less perpendicular to the plane of the sky. That was then, of course, the accepted model for all variables for a while.

The Cepheid variables, however, presented a problem. Indeed, they showed variable radial velocities with the same period as their brightness variations, though not quite with the phase relationship you would expect for an eclipse. More serious, when you went to fit an orbit to the velocity curve, the separation of the stars came out much smaller than their sizes would have to be. An alternative, radial pulsation, came first from Henry Plummer,[54,55] was expanded by Shapley,[56] and made quantitative enough to win over most of the community by Eddington.[43] Plummer is remembered, if at all, these days for models of the distribution of stars in globular clusters. Shapley, of course, got us out of the center of the Milky Way, and so Eddington normally gets the credit.

Let it be said loudly, because this is the first case in this paper where the right answer is "All of the above!". Cepheids, RR Lyraes, and other classes of variable stars pulsate, sometimes radially as Plummer and all had in mind, sometimes in much more complex modes. Eclipsing binaries are, well, eclipsing binaries. And a large number of stars that display short rotation periods and extensive spottedness have had their rotation period measured (and sometimes their spots mapped out) by the resulting variability. "Both, please" or "All of the above" is quite often the right answer to an astronomical conundrum. Portions of the answer to this one were in place soon after the 1667 and 1782 discoveries, but it didn't all fall into place until about 1926, for a duration of something like 259 years, the longest stretch in any of these sections.

3.5. *Variable stars, traumatic*

By these I mean the supernovae and classical novae (plus, if you wish, other, related cataclysmic variables). Tycho's *nova stella* of 1572 we now count among the supernovae, as is true also for Kepler's 1604 event and those of 1006 and 1054. Indeed, the first real nova to have been recorded was probably WY Sge in 1783,[57] since CK Vul of 1670 is now widely regarded as a member of a much smaller class that includes V838 Mon (with so far no very compelling models). Credit for separating out the supernovae generally goes to Walter Baade and Fritz Zwicky publishing in 1934,[58] though in fact both Lundmark[59] and H. D. Curtis as part of the Curtis–Shapley debate[60] had made a very similar distinction more than a decade earlier.

Because Baade and Zwicky in the same pair of papers cautiously advanced the idea of collapse of a normal star to yield a neutron star as the energy source, it is easy to think of the supernovae problem as having been solved at the same time it was recognized. Naturally, the situation is not quite so simple. The objects in their inventory were mostly ones that we would now call Type I supernovae (the distinction came later, from spectroscopic work by Minkowski in 1942). Just then, most astronomers were not taking either neutron stars or Zwicky very seriously, and one cannot speak of a consensus around either a mostly right or a slightly wrong idea. The second mechanism, explosive nuclear reactions, dates back at least to 1960.[61]

Today there is nearly universal agreement that the two major sorts of supernovae are core collapse (Type II; Type Ib,c) and nuclear explosion (Type Ia), but determining just how the gravitational energy from the collapse is deposited in the outer stellar layers to expel them and deciding which kinds of binary systems are most likely to drive a white dwarf to explode are both still work in progress. One cannot, therefore, really assign a duration to the total story, though I have provided a couple of snap shots[62,63] along the way.

The two main types of supernovae are now readily distinguished in terms of light curves, spectra, and remnants as well as energy source, though there is an interesting subsidiary story connected with identifying the elements responsible for the main features in SNe Ia (those with no hydrogen detectable near peak light), the tale stretching from at least 1937 (SN 1937c — we would call it SN 1937C now — in IC 4182 was very bright), when several observers declared they could not make head or tail of the spectrum, to the late 1960s, when several modelers concluded that the right answer is very broad lines of common elements, but excluding H and He.

Stars which become supernovae do not survive in anything like their original form, leaving only a neutron star or black hole + expanding gas shell (core collapse type) or just the expanding gas (nuclear explosion type). Novae, in contrast, have been known since the early 20th century[28] to settle back down to an appearance much like what preceded the event. Five naked-eye novae and many fainter ones in the first quarter of the 20th century led to very extensive studies of light curves, spectra, and expelled gas. Indeed, the 1926 Russell *et al.*[28] discussion would not badly mislead a beginning investigator today.

In contrast, the question of how the requisite large amount of energy could be released in less than a day (novae brighten very rapidly sometimes) could then be answered only with a vigorous shrugging of shoulders. Today, however, we would say that we are so confident of the answer that it at the core of our definition of novae. The whole class of cataclysmic variables (or cataclysmic binaries) is held to consist of white dwarfs in relatively tight orbits with normal stars, so that hydrogen is accreted on the white-dwarf surface. The dwarf novae and their cousins are releasing gravitational potential energy semiperiodically[64] because of instabilities in the donor star[65] or, more probably, in an accretion disk.[66] Given that the class was recognized in 1896,[67] this gives us a duration of 70-something years, though, I think, without any very popular single explanation in between.

As for the novae and recurrent novae (observationally distinguished from the dwarfs by the fact of ejected significant material), the first thing nearly everybody seems to have thought of is stellar collisions and then rejected because they would be far too rare to provide the dozen to dozens of events that must occur each year in a large spiral like the Milky Way or M31, though the idea was revived in 1939[68] as a possible mechanism for the few-per-century supernovae. What would be commoner than star–star collisions? Presumably stars being hit by planets and swarms of meteoric material. This is Russell *et al.*'s[28] best-bet hypothesis (reserving star–star events for the then-unique S Andromeda). They attribute it to Hans Seeliger and William C. Pickering, neither of whose biographers in Ref. 25 thought the idea important enough to mention. The intention was to use kinetic energy of the smash to heat layers just below the stellar surface sufficiently to enhance production of whatever sort of subatomic energy was keeping the stars in general shining (not infinitely far off the truth).

Enough other ideas were put forward that we cannot, I think, speak of any consensus in favor of planetary impacts. In 1931 Milne[69] proposed the collapse of a normal star to a white dwarf. This was instantly rendered unlikely by the similar appearance of the stars concerned before and after, but notice it meant that the idea of collapse was in the air when Baade and Zwicky went to consider supernovae. Also ignored by his BEA (Ref. 25) biographer was Biermann's[70] 1939 proposal of rapid changeover from radiative to convective energy transport in a stellar envelope. The first nuclear explanation, which required an accumulation of He^3 near a star's center, came from Schatzman in 1951.[71]

It is really ordinary hydrogen that explodes, because that is what the companion transfers. It quickly piles up to become degenerate and extra C^{12} is available from an accreting white dwarf to catalyze CNO cycle fusion. A typical text from the 1950s[72] hovers between some sort of collapse and some sort of Biermann-like atmospheric convulsion, and it was not, I think, until well into the 1960s that extended spectroscopic work by Kraft and others[73] established that the novae, dwarf novae, and all are binaries that convergence began, implying a duration of puzzlement well in excess of 40 years.

Incidentally, two classes of cataclysmic variables, the recurrent novae and the supersoft X-ray binaries, appear to burn their hydrogen and helium in sufficient peace that the mass of the underlying white dwarf gradually increases (versus having more thrown off than was accreted to make the explosion in a nova) and the system might evolve to a Type Ia supernova, showing, perhaps, that the initial single class of *novae stellae* was not so very wrong after all.

3.6. *Other stellar stories*

Are these all the tales that can be told? No, of course not, any more than Sec. 2 exhausted the solar system. For both, the most extensive unsolved problem is surely the formation mechanism, and now that it is recognized that very many stars have planets (many quite unlike ours) and that our own system consists entirely of objects that are all the same age (5.5 Gyr roughly), the two problems are beginning to coalesce. For each, there was an earlier stage of extended erroneous consensus, however. For the solar system, this was the period from about 1890 into the 1940s when the "best buy" hypothesis was the Chamberlin–Moulton proposal that a passing star had pulled out solar material to become the planets. This gradually gave way to a rebirth of the Kant–LaPlace hypothesis of co-formation from a disk.

In the case of star formation, the curious period was one from the late 1920s until some years after WWII, when its very existence as an ongoing phenomenon was denied, and the formation of both stars and galaxies relegated to "long ago, when conditions were very different." I have told much of that story elsewhere.[74]

The temporal evolution of stellar rotation rates and activity levels, the causes of magnetic cycles (in both Earth and Sun!), the processes that accelerate a few particles to very high energies in supernovae and elsewhere, and many other topics are unfinished stories, though again some of them with long stopovers at later-discredited answers.

4. Milky Way Mysteries

4.1. *The location of the Sun in the galaxy*

Copernicus got us out of the center of the universe sometime between the 1543 publication of his *De Revolutionibus* and the 1609 publication of Kepler's *Astronomia Nova.*[75] In the same period and for a couple of centuries beyond, philosophers (both natural and artificial) put forward a wide range of schemes, in some of which the solar system had a central position and in others of which it did not. Classification and lots of pictures appear in two volumes by Edward R. Harrison.[35,76] It was William Herschel who stuck us firmly in the middle in 1785.[77] His argument was that in counting stars with his 20-foot (the focal length, not the mirror diameter!) telescope, he ran out of stars at roughly the same apparent brightness, hence at the same distance if all stars are like the sun, in all directions in a great circle, the Milky Way, around the sky. He ran out more quickly perpendicular to that great circle, yielding a disk of stars.

When Herschel's 40-foot telescope revealed fainter stars, he realized that nothing could be said about the size or shape of the galaxy by these methods.[78] Curiously, the us-centered disk drawing has continued to be reproduced as the definitive word of the master down to the present time. Over the years, assorted pundits drew assorted conclusions; Alfred Russel Wallace, for instance, declared that no other place in the universe than this central region would or could be habitable,[79] in accordance with a sort of anthropic argument. Darwin, in contrast, appears to have supported a plurality of worlds. Wallace also rejected life on Mars.

Despite occasional alternative suggestions (a particularly charming one has us at the center of a circular galactic disk, but spiral arms centered far away in Cygnus),[80] it took the 60″ telescope on Mt. Wilson and half a decade of hard work by Harlow Shapley to pry us out of the galactic center.[81] Shapley concentrated on pulsating variable stars (he was among the first to recognize that Cepheids work that way) in the globular clusters of the Milky Way and developed a distance scale for them. When he plotted cluster locations, they formed a more or less spherical distribution centered in the direction of Sagittarius and something like 18–20 kpc away (the modern value is 7.5–8.5 kpc).

Kapteyn's definitive Sun-centered galaxy[82] actually appeared after Shapley had completed his globular cluster studies and indeed cited them as demonstrating the near-complete transparency of interstellar space (next section). A good many astronomers attempted to visualize a Milky Way or universe in which both could be correct. Russell *et al.*[28] and Robert Trumpler,[83] among others, attempted to place the Kapteyn universe as a 2–3-kpc-wide concentration of stars, located in the plane of Shapley's larger Milky Way and centered somewhere near the sun. Gould's belt of B stars makes up part of that concentration and is a real feature.

If you date the recognition that we must live somewhere in the Milky Way to Herschel's 1785 drawing and the definitive solution to a few years after Shapley's papers, then the duration was more than 135 years, during nearly all of which we lived essentially at the center, even in such enlightened eyes as those of Eddington in 1912.[84] In retrospect, Shapley's concentration on clusters far from the plane of the Milky Way guaranteed that the assumption of zero absorption (based on an absence of detectable reddening for even the most distant) would seem to be verified. The error, however, propagated for decades and was one of several major contributing factors in the prolonged difficulty in determining accurate distance scales outside the Milky Way.

4.2. *The transparency of space*

That distant lights look faint was presumably discovered by the first paleolithic tribe to carry its fires around. That the drop-off should be a $1/d^2$ Coulomb's law as for gravity seems to have been accepted by Descartes, Newton, and other contemporaries. Indeed, the idea can be traced back to one of the Greek philosophical lines of thought about the directness of sense perceptions. All, of course, accepted that the Earth's atmosphere could block light as well as bending it and the 1340-something discussion about bending[85] (with air thinning with height and gradually giving way to the noninterfering fire element) is an interesting one.

The first explicit discussion of imperfectly transparent space came in 1744, from J. P. Loys de Cheseaux.[35,76] He had recognized the riddle that is now generally called Olbers's paradox, and said you could keep the sky background down to a reasonable level in an infinite, homogeneous distribution of stars if space is merely 330,000 trillion times (3.3×10^{17}) more transparent than water. If you can see 100 yards (or meters) through very clear water, then the attenuation length for "empty space" is 3.3×10^{19} m or 1000 parsecs.

In traditional astronomers' units and arithmetic, where $2 = e = (100)^{1/5}$ (the luminosity ratio corresponding to a brightness difference of one magnitude), this is, obviously, absorption at the rate of one magnitude per kiloparsec. I cannot remember seeing this number any place before the envelope back on which I wrote it out during a particularly uncompelling faculty meeting. Such non-transparency will put an edge to the visible galaxy at a kpc or two, much as Herschel later found.

Olbers, writing in 1823[35,76] after the publication of Herschel's heliocentric galaxy phrased the absorption he wanted as losing one ray in 800 on the way from

Sirius, whose distance from us is about 1.3 pc, so we lose half the rays in 500 parsecs, remarkably close to both Cheseaux's number and the current one. The Cheseaux–Olbers argument was, of course, a fallacious one: whatever absorbs light will eventually heat up until it radiates as much as it absorbs, but this had not been established in their era, and so they were able to conclude that they had darkened the night sky.

The topic came into general discussion in the second half of the 19th century and at the beginning of the 20th. F. G. W. Struve deduced 1 mag/kpc in 1847[25] and other estimates came from Turner, Comstock, and Halm. Indeed, Kapteyn himself suggested[86] 1.6 or 0.3 mag/kpc in 1904 and 1909, but Kienle[25] and others claimed much smaller upper limits to general interstellar absorption. Part of the problem was that it was understood that reddening should accompany scattering or absorption of white light (consider the setting Sun!), and that the effects of observing different stars (etc.) at different heights above the horizon (often correlated with location in the galaxy for practical observing reasons) had to be subtracted. Photography did not much improve the situation, because scattering of blue light in the emulsion could be another confounding source of reddening.

Shapley (previous section) made things worse, because indeed there is almost no absorption on most of the sight lines to high latitude globular clusters, and by assuming there was also none in the galactic plane he could tie his cluster Cepheids into the same period–luminosity relation as for those in the plane (now known to be a couple of magnitudes brighter).

Who was the last person to get it wrong? "Kapteyn!" will shout astronomers who know a bit (but perhaps only a bit) of their own history. Well, yes and no.[86] He must surely have counted more stars than anyone before him in the 30 years centered around 1900, and he knew, of course, about Barnard's[25] opaque but localized clouds. As he sat down to write his most extensive paper[82] on the distribution of stars in space, he opted for zero diffuse absorption. Partly, he said, this was influenced by Shapley's conclusion that the globular clusters were not reddened and partly, perhaps, by his recognition that he did not have much time to complete the work and was still in no position to deduce a definitive absorption rate. He died that same year (1922).

The community thus soldiered on with transparent space and Shapley's brightness scale for Cepheids, making the Milky Way look a good deal bigger and brighter than it really is — a disincentive to take the possibility of existence of other galaxies seriously, particularly in Shapley's mind, while a smaller Herschel–Kapteyn Milky Way allowed Heber Doust Curtis to accept them.[87] And when other galaxies forced themselves to the fore (next section), the incorrect Cepheid scale put them far too close and so apparently much smaller and fainter than the Milky Way.

As the depression approached, there were several near-misses on discovery of interstellar absorption and its implications, including Jesse L. Greenstein at Harvard,[88] who backed off in favor of atmospheric effects, and Carl Schalen at Lund,[25] who published 0.5 mag/kpc in 1929. Nevertheless, astronomical folklore gives just about 100% of the credit to Robert J. Trumpler[83] publishing from Lick the next year. He refrained from trying to count stars yet again and instead examined the apparent angular diameters versus the apparent brightnesses of young (galactic rather than globular) star clusters. The correlation was not the expected quadratic. Instead, the smaller angular diameters went with excess faintness, as if their light was being absorbed at a rate of about 1 mag/kpc, and the best value has hovered there ever since.

It is a little difficult to assign a duration to the majority view of empty, transparent space. Perhaps Herschel to Schalen or only Shapley to Trumpler, a mere dozen years, but it was still 1944 before Henri Mineur folded Trumpler's numbers[25] into the Cepheid brightness scale. And, as there was a war going on, very few noticed, so that Baade was surprised in 1948 when he turned the new 200-inch telescope toward M31 and saw no RR Lyraes there, doubling the extragalactic distance scale for the first of several times.[89,90]

4.3. *The stability of spiral arms*

In the 1850s, William Parsons[25] (aka Third Earl of Rosse) established the spiral appearance of a number of nebulae, now known to be external galaxies. Suppose that the material in the arms is on Keplerian orbits around a central mass. Then the arms will surely stretch out and wind up after a few rotation periods. This was regarded as good by the first person known to have remarked upon it in print, E. J. Wilczynski in 1896,[91] because he was trying to account for the formation of spiral shapes. That the spiral nebulae indeed rotate was established spectroscopically by Wolf, Slipher,[28] and Pease[25] before the end of WWII, and the velocities were hundreds of km/s.

In the same time frame, Adriaan van Maanen thought he had seen rotation of spirals in the plane of the sky (requiring them to be within the Milky Way), and I will say no more about this here except to remark that just what went wrong is sufficiently contentious that the author (Adriaan Blaauw) and the editor (me) of the van Maanen article in Ref. 25 were able to remain friends only because Blaauw is an extraordinary gentleman.

Establishing that the Milky Way rotates and is a spiral was a good deal more difficult (for some of the same reasons that very few fish are ichthyologists), and the details were established only with the advent of radio astronomy. But the optical evidence was reasonably persuasive to Bertil Lindblad[25] and confirmed in 1927 by Jan Oort.[25] And it was Lindblad who worried about stability of spirals from at least 1941 until shortly before his death in 1965[92] and proposed a form of stable wave behavior called a soliton (known in water waves, condensed matter behavior, and various other contexts) which can propagate at a speed different from the fluid in which it moves and maintain its shape and amplitude with little dissipation over long times. Lots of the difficult mathematics was done by Lin and Shu,[93] and there has been general consensus since that this must be at least part of the answer for

Grand Design spirals. Other, less-organized-looking arms may well form, stretch out, and dissipate many times over the life of a disk galaxy.

Probably no one would claim that this territory has been fully mapped, but if "Oh, they're density waves" is enough of an answer for your purposes, then the duration of the puzzle is the 68 years from 1896 to Lin and Shu's 1964 paper. Bars, other instabilities, and interactions with nearby galaxies are surely also part of the picture.

4.4. *More of the Milky Way*

Our galaxy is apparently a typical spiral, so that many of the things that puzzle us about it — numbers and types of companion galaxies and stellar populations, how it all got put together, nature of the dark matter — belong in Sec. 5 as stories about galaxies and the universe in general, many of which do not yet have a *denouement.*

Interstellar polarization was a surprise when discovered — the observers had been looking for something intrinsic to B stars[94,95] — but was explained quickly enough[96,97] that there wasn't much time for puzzlement, though possibly neither of the first two mechanisms contains the entire truth.

5. Cosmic Queries

Very many of these — like the nature of dark matter and dark energy, and what came before the big bang — are ongoing. On the other hand, the existence of dark matter and the occurrence of a big bang (meaning a hot, dense stage about 14 Gyr ago) are reasonably well established after extended periods of dispute, and so are legitimate topics for our consideration here.

5.1. *The existence of other galaxies*

The ancient Greeks left us a choice of three relevant ideas.[35] The universe of Aristotle was unique and finite. Einstein's solution to his own equations was initially of that sort — as well, of course, as static. The Stoics had a single universe with an infinite void outside, and the Epicureans an infinite number of worlds, both like and unlike ours, spread through infinite space. These latter two correspond at least roughly to the two competing ideas from the 1700s down to about 1925 of the Milky Way as the universe and of "island universes" or multiple external galaxies.

The geocentric universe of Aquinas (*c.* 1350) and the heliocentric one of Copernicus (published in 1543) were both firmly Aristotelian in the sense of being finite and the whole show. Thomas Digges (1576) and Giordano Bruno (before 1600!) put forward infinite universes, with many worlds. But Kepler in 1606, for instance,[35] was absolutely horrified by the idea of infinity and its implications (not uncommon down to the present).

We then skip forward another 200 years to 1750–1760, when Thomas Wright, Immanuel Kant, and J. H. Lambert[25] concurred that the Milky Way is a disk of stars, others of which might have planets and inhabitants. Lambert voted for a finite system and debated whether the nebulae then known were other Milky Ways. Kant said they were, while Wright's highest-order structure was a sphere with gravitational and moral center coinciding, and that he turns up as often as he does in introductory astronomy texts appears to have resulted from Kant giving him more credit than he deserved.[25] Notice that all three of the Greek alternatives are still represented here, as they will be down nearly to the present.

If you feel you have been deprived of the 1600s, well you have, because I find the attempts by Newton and others to invoke equations and divinity in the same stroke of the pen heavy going. Suffice it to say that Descartes was, in the present context, an Epicurean and Newton a Stoic, though neither took a stand on the nature of the nebulae, very few of which were known at the time.

The William Herschel of 1785 was a "nebulae are Milky Ways" supporter, and it was in that context that he remarked that he had seen light that left its stars millions of years ago. But by 1811, he had arranged the nebulae into an evolutionary sequence connected with stars and star clusters and decided that the Milky Way was the universe, though he no longer thought he could see to its edges. Humboldt was a many-galaxies fellow, who is said to have coined the name "Island Universes" in about 1855,[35] in the German equivalent of "cosmic islands."

It might have seemed at this point as if the Epicureans were winning, but Huggins's demonstration in the 1860s that many nebulae consisted of hot, tenuous gas[35] began to swing the pendulum back to a single stellar ensemble with nebulae in and around it and empty space beyond. This is the only structure suggested by Simon Newcomb in his *Popular Astronomy* (1878). Other items seemingly favoring the single-galaxy universe were the extreme brightness of S Andromeda in 1885 (now known to have been a supernova, but then thought likely to be a star that had ventured into a nebula), the Stoic solution to Olbers's riddle, and (erroneous) measurements of proper motions within M31 and M51 in 1899 by Isaac Roberts.[35]

The science popularizer Agnes Mary Clerke is almost universally quoted at this point,[25,35,26] because she phrased it so nastily: "No competent thinker, with the whole of the available evidence before him, can now, it is safe to say, maintain any single nebula to be a star system of coordinate rank with the Milky Way." She died in 1907, just before the pendulum began to swing back again, with Carl Charlier's[25] 1908 argument that the darkness of the night sky could be explained by a hierarchical distribution of nebulae (which indeed he, and you, can see by plotting out the objects in Dreyer's *New General Catalogue*).

Over the next 15 years, Charlier and H. D. Curtis argued that nebulae avoid the plane of the Milky Way because of obscuring dust clouds there; V. M. Slipher began to find very large radial velocities for spiral nebulae; and Ernst Öpik actually estimated a 450,000 pc distance to Andromeda by assuming a stellar population like the Milky Way and Slipher's rotation speed for it.

On the other side, Shapley held to his single, enormous Milky Way; Kapteyn put us at the center of a much smaller, but unique, one; Keeler thought[25] the spectra he was getting at Lick of spirals acted like solar systems in formation; and, probably most important, van Maanen reported proper motions in several nebulae. These were in the direction of knots moving outward on the spiral arms — that is, leading, rather than trailing, arms in opposition to the radial velocity evidence.

The definitive word was written by Edwin Powell Hubble, about whom whole books have been written.[98] But enough of the story is told in any standard reference. Between 1923 and 1925 he found Cepheid variables, first in NGC 6822, and (first published and announced) in M31. The distance scale, though admittedly too small, was good enough to say that these must be separate galaxies, though, if they were islands, the Milky Way was, according to Shapley, more like a continent. The issue was regarded as already completely settled by Russell *et al.*[28]

With high-grade hindsight, many astronomers have said that the issue should not have been in doubt after the demonstration that first M31 (Scheiner, 1899) and then other spirals (Fath, 1909; Sanford, 1917)[25] had absorption line spectra like the sun and star clusters. The era of confusion and partial erroneous convergence can be said to extend, if you wish, from ancient Greek times to 1925, or perhaps only from the mid-19th-century to 1925.

5.2. *The cosmic distance scale and age*

We have already noted (Sec. 4.2) that Shapley's distance scale for Cepheid variables made the Milky Way appear larger than it is and other galaxies closer, when Edwin Hubble came to set a cosmic distance scale in 1929. That scale plus the redshifts found first by Slipher and later by Milton Lassell Humason[25] yielded a cosmic expansion rate of 536 km/s/Mpc[99] with, according to Hubble, about 10% (statistical) uncertainty. That 10% has actually persisted down to the present day. Over the next couple of decades, Hubble suggested[98] other values between 500 and 550 for what we now call the Hubble constant, implying cosmic ages somewhat less than 2 Gyr, that is $t \sim 1/\mathrm{H}$.

Hubble's linear velocity–distance relation meshed so neatly with expanding solutions of Einstein's equations that expansion of the universe was a majority view very quickly, though Zwicky[100] put forward an alternative, tired light, model also in 1929, and Hubble, not in general a theorist, was never as firm about expansion as you might have expected for the discoverer.[98] And, naturally, everybody used Hubble's numbers, for more than 20 years.

Notice that this is not really the same sort of beast as community consensus around a wrong idea. Indeed, if observers want to be able to compare their results for different galaxies, it is arguably much more important that they all use the same value of H and distance scale than that it be the right one. Post-Shapley views on our distance to the galactic center come under the same heading, and it was therefore not foolish that the galactic research community engaged in formal votes on the issue and agreed to use a common number.[101]

The best-buy value of H started coming down in 1952 and had taken enough steps — all sensible and all predicated on a better understanding of what was being observed and how to deal with statistical issues — that, by 1965. Caltech graduate students were betting that H might go negative, and the universe start to contract.[89,90] This did not happen. Rather, two schools of thought hardened their positions and faced off defending $H = 100$ and $H = 55$, and it became common to use h, meaning H in units of 100 km/s/Mpc. This still sometimes happens, though h_{70}, meaning H in units of 70 km/s/Mpc, has become at least as common.

One of (many) things the Hubble Space Telescope was supposed to do was provide a definitive value for H. This actually happened, and the report of the Key Project Team[102] that $H = 72 \pm 8$ (that faithful 10% uncertainty) is now very widely used, at least to the extent that anyone holding out for a value less than 65 or more than 75 will make very few friends. The HST report was the most heavily cited astronomy paper of 2001.

Use of earlier, larger (or, occasionally, smaller) values of H lasted from 5 to 25 years each, and it is important to figure out which was employed when trying, for instance to decide how much dark matter Zwicky was claiming in the Coma cluster in 1933.

5.3. *The existence of dark matter*

The first thing to be said is that accepting "existence" is not at all the same as knowing what the stuff might be ("nature" of dark matter), and the second is that convergence has not quite occurred, because the alternative of a theory of gravity in which G gets a bit stronger at large distances still has adherents on active duty, in contrast to the "big bang" situation below.

I have not attempted to determine what various ancient Greeks might have thought about matter that could not be seen, but real, physical, gravitating dark objects were the first explanation for Algol-type variability put forward by Goodricke and Pigott (Sec. 2.1), and Pigott[25] suggested that nebulae like the Coal Sack might be populated by dark stars. If two is a confirmation, then dark astronomical objects were already established in the 1780s, since John Michell[25] proposed that sufficiently massive objects could have their light return to them "by their own proper gravity."

You may then have 150 or so years off until Jeans and Kapteyn looked for dark stars or dark matter (yes, the word goes back that far) in 1922[103] and Oort in 1932, and Zwicky reported "*dunkle materie*" in the Coma cluster. I think, however, that the recognition of the concept as an interesting one should be dated to a 1961 conference[104] that was convened specifically to discuss the meaning of large velocity dispersions in clusters of galaxies. Though definitive evidence that mass-to-light ratios of astronomical systems increase on longer length scales could have been put together before WWII, drawing on the work of Hubble, Zwicky, Smith, Babcock, and Holmberg,[105] nobody seemed to think this very important.

In contrast, at the 1961 meeting, whose proceedings are fairly frank by current standards, there were respected voices favoring dark matter, suggesting that the clusters were actually unstable, and warning of large observational uncertainties. The first $G = G(r)$ model followed the next year.[106]

Such models have become very much more sophisticated in the intervening 45 years, but somehow spend most of their time (or anyhow the time of their proponents) chasing after successive objections coming from the need to account for gravitational lensing and the details of the cosmic microwave background as well as rotation curves of galaxies and velocity dispersions in clusters.

If you want a date for when dark matter (of whatever type) probably became the dominant opinion, 1974 would be a good choice because of a pair of short papers[107,108] that did precisely what I suggested a couple of paragraphs ago. The authors plotted or tabulated mass-to-light ratios from the size scale of star clusters and galactic disks on up to superclusters of galaxies and found a monotonic increase. There has not (yet!) been any convergence around any other idea, so that you could set a duration of only 13 years (1974 minus 1961) to the duration of puzzlement. Or, if you are one of the alternative-theory-of-gravity fans, you would say that the community has, so far, been collectively wrong from 1961 to 2008 and still awaits enlightenment (or, perhaps, endarkenment) from them.

5.4. *The standard hot big bang*

Once again, you might quibble with the inclusion of this topic. First, there is a whole book about it,[109] to which obviously I can add very little. Second, convergence is not (yet) complete, though in contrast to the dark matter situation, the opposing camp now includes almost exclusively astronomers (etc.) past normal retirement age, and indeed the most vocal, Fred Hoyle, died in 2001.

There is, however, a more gentle sort of opposition, coming from (frequently) physicists who do not like the idea of time = 0 and density = infinity. The same sort of objection, often by the same objectors, has been raised to the concept of black holes. But an astrophysicist's black hole is simply something of a size not much larger than the Schwarzschild radius for its mass and often with some of the redshifting and frame-dragging properties of an external Schwarzschild or Kerr solution. The observational evidence for these is very strong.

Similarly, very many lines of observational evidence converge on the idea that, about 14 Gyr ago, the material that forms our observable universe passed through a stage of very high temperature and density, close to thermal equilibrium and homogeneity. The evidence includes the existence of the cosmic microwave background and many of its details,[110] the abundances of deuterium, helium, and lithium-7 in the most nearly unprocessed material still to be found, ages of the oldest stars, very different appearances of the most distant galaxies (etc.) from those here and now, apparent brightnesses of distant supernovae, and properties of the largest-scale structures to be found today. Together these are the underpinnings of a standard cosmological model, which includes the "proper" value of H, a certain amount of dark matter, and also a dominant component called dark energy, about which I will say no more here. This might be a good moment to recall that Kelvin's confidence in his age for the Sun and for the Earth was strengthened by the two numbers agreeing. And back when everyone knew we were at the center of the Milky Way, old (by then) Simon Newcomb asked whether this might not be the same sort of delusion that Ptolemy had suffered from.[31]

The main opposition to a hot big bang from 1948 until the early 1960s, was, of course, the steady state cosmology proposed by Hermann Bondi and Thomas Gold and by Fred Hoyle. This made very definite predictions and has clearly now been falsified.[109] What survives among perhaps a double handful of former steady state proponents and adherents of non-cosmological redshifts is more like a feeling of disliking many aspects of the so-called standard model and a tendency to warn that we are all going in the wrong direction and will eventually walk off the edge of the universe (symbolically of course).

If, on the other hand, you would like a hot topic to work on, then the operational definition of big bang just given allows you to ask what came before and to explore answers like inflation, strings, and branes. If you allow the BB versus SS cosmological controversy a lifetime of 15–20 years, then perhaps "what came before" could be sorted out by 2028, but I won't bet money on it.

Please be clear that there was never convergence of majority opinion on anything except a relativistic, evolutionary universe. In the immortal words of Allan Sandage, "I think it's true to say that no one in southern California ever took steady-state seriously" (p. 285 of Ref. 109). He must have meant Mt. Wilson-Palomar Observatories by "southern California," since there were folks who took it quite seriously at both the California Institute of Technology and the University of California, San Diego. Perhaps there still are.

6. Applications in Other Fields?

All these stories (and I have a long list not included here) belong to the realm of astronomy. Are there similar ones in other fields? Undoubtedly. Time was when there was a 1460-year disagreement about the dating of the earliest Egyptian dynasties, the 1460-year period being the time it takes for a correct solar year to lap a 365-day one once (Sothos cycle), that went back to the 1920s or earlier. I entered the fray[112] with an imprecise, astronomical method (pyramid alignments) shortly after convergence had occurred in the early 1960s via other more traditional archeological approaches, luckily on the right side.

Current controversies in areas where there had (perhaps) been previous agreement include when and by what route(s) *Homo sapiens* entered the western hemisphere[111]; whether agriculture arose in more than one place independently (probably yes); whether writing arose in more than one place independently (possibly no?); and how much, if any, genetic interchange was there between *H. sapiens* and *H. neanderthalis*. I hope a suitably skilled anthropological storyteller writes up some of these one day!

Acknowledgments

I am grateful to Drs. Phua and Tan for this opportunity to collect and augment some of the astronomical tales of "*il etait une fois...*" from the history of astronomy. It was Thomas Hockey who offered me the opportunity to be one of the senior editors (and authors) of the *Biographical Encyclopedia of Astronomers.* This multiyear task, perhaps like reading the complete works of Charles Dickens, is the sort of thing one is very glad to have done after the fact, though perhaps with reservations in between. Historians quoted frequently here whose views have strongly influenced mine (without in any sense suggesting that they would agree with either details or the general approach) include David DeVorkin, Michael Hoskin, Dennis Danielson, Alan Hirshfeld, Owen Gingerich, and Helge Kragh.

References

1. R. W. Klebesadel, I. B. Strong and R. A. Olsen, *Astrophys. J.* **182** (1973) L85.
2. S. G. Djorgovski *et al.*, *Nature* **387** (1997) 876.
3. M. R. Metzger *et al.*, *Nature* **387** (1997) 878.
4. A. Hewish *et al.*, *Nature* **217** (1968) 709.
5. V. I. Ginzburg, *Dokl. Akad. Nauk.* **156** (1964) 43 (in Russian).
6. L. Woltjer, *Astrophys. J.* **140** (1964) 1309.
7. F. Pacini, *Nature* **216** (1967) 567.
8. F. Pacini, *Nature* **219** (1968) 145.
9. T. Gold, *Nature* **218** (1968) 731.
10. R. F. Christy, unpublished calculations, performed at the Institute of Theoretical Astronomy, Cambridge, UK (Spring, 1968).
11. R. J. Nemiroff, *Comm. Astrophys.* **17** (1994) 189.
12. V. Trimble in S. S. Holt *et al.* (eds.), *Gamma Ray Bursts in the SWIFT Era, AIP Conf. Proc.* **836** (2006) 3.
13. W. Baade and F. Zwicky, *Proc, U.S. Natl. Acad. Sci.* **20** (1934) 254, 259.
14. E. P. Hubble, *Proc. U.S. Natl. Acad. Sci.* **15** (1929) 168.
15. W. Baade, in *Trans. IAU*, ed. P. Th. Oosterhof, **8** (1954) 297.
16. M. L. Humason, N. U. Mayall and A. Sandage, *Astron. J.* **61** (1956) 97.
17. A. Sandage, in *Problems of Extragalactic Research*, ed. G. C. McVittie (Macmillan, New York, 1962), p. 359.
18. A. Sandage and G. A. Tammann, *Astrophys. J.* **210** (1976) 7.
19. W. L. Freedman *et al.*, *Astrophys. J.* **553** (2001) 47.
20. A. Wolszczan and D. A. Frail, *Nature* **355** (1992) 145.
21. M. Mayor and D. Queloz, *Nature* **378** (1995) 357.
22. G. W. Marcy and R. P. Butler, *Astrophys. J.* **464** (1996) L147.
23. A. W. Hirshfeld, *Parallax: The Race to Measure the Cosmos* (W. H. Freeman, 2001).
24. R. Baum and W. Sheehan, *In Search of Planet Vulcan: The Ghost in Newton's Clockwork Universe* (Plenum, New York, 1997).
25. T. Hockey *et al.* (eds.), *The Biographical Encyclopedia of Astronomers* (Springer, 2007).
26. M. Hoskin (ed.) *Cambridge Illustrated History of Astronomy* (Cambridge University Press, 1997).
27. R. A. Lyttleton, personal communication, 1969.
28. H. N. Russell, R. S. Dugan and J. Q. Stewart, *Astronomy* (Ginn, Boston, 1926).

29. T. Hockey, *Galileo's Planet: Observing Jupiter Before Photography* (Institute of Physics, London, 1999).
30. U.-J.-J. Le Verrier, *Ann. Obs. Paris* (1859) 5.
31. S. Newcomb, *Sidelights on Astronomy and Kindred Fields of Popular Science* (Harper, New York, 1906).
32. A. Einstein, *Sitzungsber. Preuss. Akad. Wiss* (1915) 831.
33. Ralph Baldwin, personal communication, *c.* 1995.
34. C. Brans and R. H. Dicke, *Phys. Rev.* **124** (1961) 925.
35. E. R. Harrison, *Cosmology, the Science of the Universe*, 2nd edn. (Cambridge University Press, 2000).
36. V. Trimble, *BeamLine* **24**(1) (1994) 35.
37. S. Baxter, *Ages in Chaos* (Forge, New York, 2003).
38. D. Lindley, *Degrees Kelvin* (Joseph Henry Press, Washington DC, 2004).
39. F. P. Venable, *Chem. News* **59** (1889) 89.
40. E. R. Scerri, *The Periodic Table* (Oxford University Press, 2007).
41. C. H. Payne, *Stellar Atmospheres* (Heffer & Sons, Cambridge, 1925).
42. C. H. Payne, *Washington Natl. Acad. Proc.* **11** (1925) 192.
43. A. S. Eddington, *The Internal Constitution of the Stars* (Cambridge University Press, 1926).
44. R. d'E. Atkinson and F. Houtermans, *Z. Phys.* **54** (1929) 656.
45. J. N. Bahcall and E. E. Salpeter, in *Hans Bethe and His Physics*, eds. G. D. Brown and C.-H. Lee (World Scientific, Singapore, 2006), p. 147.
46. M. Schwarzschild, *Structure and Evolution of the Stars* (Princeton University Press, 1958).
47. J. N. Bahcall, *Neutrino Astrophysics* (Cambridge University Press, 1989).
48. J. N. Bahcall and H. A. Bethe, *Phys. Rev. Lett.* **65** (1990) 2233.
49. W. Crookes, *Chem. News* **60** (1889) 27, 3 following articles.
50. J. W. Nicholson, *Mon. Not. R. Astron. Soc.* **74** (1917) 506.
51. S. Arrhenius, *Z. Elektrochem.* **28** (1922) 405.
52. P. Swings, *Astron. J.* **98** (1943) 119.
53. P. Moore, *Astronomers' Stars* (W. W. Norton, New York, 1989), Chaps. 7, 10.
54. H. C. Plummer, *Mon. Not. R. Astron. Soc.* **73** (1912) 665.
55. H. C. Plummer, *Mon. Not. R. Astron. Soc.* **74** (1913) 662.
56. H. Shapley, *Astrophys. J.* **40** (1914) 448.
57. C. H. Payne Gaposchkin, *The Galactic Novae* (North-Holland, Amsterdem, 1957).
58. W. Baade and F. Zwicky, *Proc. U.S. Natl. Acad. Sci.* **20** (1934) 254, 259.
59. K. Lundmark, *Sven. Vetekapskad, Handlingar* **60** (1920) No. 8.
60. H. D. Curtis and H. Shapley, *Bull. Natl. Res. Council* (*U.S.*) **2** (1921) Part 3, 171.
61. F. Hoyle and W. A. Fowler, *Astrophys. J.* **132** (1960) 565.
62. V. Trimble, *Rev. Mod. Phys.* **54** (1982) 1183.
63. V. Trimble, in *Texas in Tuscany*, eds. R. Bandiera *et al.* (World Scientific, Singapore, 2003), p. 269.
64. J. A. Crawford and R. P. Kraft, *Astrophys. J.* **123** (1956) 44.
65. B. Paczyński, *Acta Astron.* **15** (1968) 89.
66. Y. Osaki, *Publ. Astron. Soc. Japan* **26** (2974) 429.
67. L. D. Wells, Harvard College Observatory Circular (1896) No. 12.
68. F. Whipple, *Proc. U.S. Natl. Acad. Sci.* **25** (1939) 118.
69. E. A. Milne, *Observatory* **54** (1931) 144.
70. L. Biermann, *Z. Astrophys.* **280** (1939) 344.
71. E. Schatzmann, *Ann. d'Ap.* **14** (1951) 294.

72. L. H. Aller, *Astrophysics: Nuclear Transformations, Stellar Interiors, and Nebulae* (Ronald, New York, 1954), p. 177.
73. R. P. Kraft, *Adv. Astron. Astrophys.* **2** (1963) 43.
74. V. Trimble, in *Star Formation Near and Far*, eds. S. B. Holt and L. G. Mundy (AIP Conference Series 393, New York, 1997), p. 15.
75. D. Danielson, *The First Copernican: George Joachim Rheticus and the Rise of the Copernican Revolution* (Walker, New York, 2006).
76. E. R. Harrison, *Darkness at Night* (Harvard University Press, Cambridge, 1987).
77. W. Herschel, *Philos. Trans. R. Soc. (London)* **75** (1785) 213.
78. M. A. Hoskin, *The Herschels of Hanover* (Science History, Cambridge, 2007).
79. A. R. Wallace, *Man's Place in the Universe* (Chapman & Hall, London, 1903).
80. C. Easton, *Astrophys. J.* **12** (1900) 136.
81. H. Shapley, *Astrophys. J.* **49** (1919) 311, and six previous papers in the series.
82. J. C. Kapteyn, *Astrophys. J.* **55** (1922) 302.
83. R. J. Trumpler, *Lick Observatory Bull.* **14** (1930) 154.
84. A. S. Eddington, *Stellar Movements* (Macmillan, London, 1914), p. 31.
85. D. Burton, *Nicole Oresme's De visone stellarum (On Seeing Stars)* (Brill, Leiden, 2006).
86. R. W. Smith, in *The Legacy of J. C. Kapteyn*, eds. P. C. van der Kruit and K. van Berker (Kluwer, Dordrecht, 2000), Chap. 8.
87. V. Trimble, *Publ. Astron. Soc. Pacific* **107** (1995) 1133.
88. J. L. Greenstein, personal communication, *C.* 1990.
89. V. Trimble, *Publ. Astron. Soc. Pacific* **108** (1996) 1073.
90. V. Trimble, *Space Sci. Rev.* **79** (1997) 793.
91. E. J. Wilczynski, *Astrophys. J.* **4** (1896) 97.
92. A. Toomre, *Ann. Rev. Astron. Astrophys.* **15** (1977) 437.
93. C. C. Lin and F. H. Shu, *Astrophys. J.* **130** (1964) 646.
94. J. S. Hall, *Science* **109** (1949) 166.
95. W. A Hiltner, *Science* **109** (1949) 165.
96. L. Davis and J. L. Greenstein, *Astrophys. J.* **114** (1951) 206.
97. T. Gold, *Nature* **169** (1952) 362.
98. G. E. Christianson, *Edwin Hubble: Mariner of the Nebulae* (University of Chicago Press, 1996).
99. E. P. Hubble, *Proc. U.S. Natl. Acad. Sci.* **15** (1929) 167.
100. F. Zwicky, *Proc. U.S. Natl. Acad. Sci.* **15** (1929) 726.
101. V. Trimble, *Commun. Astrophys.* **11** (1986) 257.
102. W. L. Freedman *et al.*, *Astrophys. J.* **553** (2001) 47.
103. V. Trimble, *Ann. Rev. Astron. Astrophys.* **35** (1987) 425.
104. J. Neyman, T. Page and E. Scott (eds.), *Astron. J.* **66** (1961) 66.
105. V. Trimble, in *Neutrinos and Explosive Events*, eds. M. M. Shapiro *et al.*, NATO Science Series II-209 (Springer, 2005), p. 181.
106. A. Finzi, *Mon. Not. R. Astron. Soc.* **127** (1963) 21.
107. J. P. Ostriker *et al.*, *Astrophys. J.* **193** (1974) L1.
108. J. Einasto *et al.*, *Nature* **250** (1974) 309.
109. H. Kragh, *Cosmology and Controversy* (Princeton University Press, 1996).
110. D. N. Spergel *et al.*, *Astrophys. J. Suppl.* **170** (2007) 377.
111. T. Goebel *et al.*, *Science* **319** (2008) 1497.
112. V. Trimble, *Mitt. Inst. Orientforschung* **10** (1964) 183.
113. I. De Moortel *et al.*, *Astron. Geophys.* **49** (2008) 3.21.