History of Dark Matter in Galaxies

This is a preprint of a chapter which will appear in
Galactic Structure, vol 5 of Planets, Stars and Stellar Systems.
Volume editor : Gerry Gilmore, series editor Terry Oswalt, publ: Springer

Virginia Trimble
Department of Physics and Astronomy
University o California
Irivne, CA 92697 USA

Abstract
The phrase dark matter goes back to 1922, and the concept of things that exist but are not, or cannot be, seen to the ancients. Here is one version of how the phrase was gradually restricted to kinds of matter not capable of electromagnetic interactions, how evidence for its existence very gradually accumulated, and how a number of watershed events have brought the astronomical community to near consensus that there is dark matter on many scales, but that its nature remains uncertain. The acceptance of dark matter in the universe has been described as a paradigm shift but also has many of the characteristics of normal science, in which data are acquired in response to existing ideas and those ideas gradually modified in light of the data. The approach is largely chronological, but with frequent looks ahead to see how various parts of the story turn out. It is not claimed that present understanding is complete or final.

Key words: dark matter, history of astronomy, cosmology

I. Introduction: Paleomatter and the Nature of History of Science

That what we get is sometimes more than what we see is a very old idea in Western culture. The counter-earth of the Pythagorean Philolaus (4th century BCE) was postulated to be forever undetectable. Kragh (2007, p. 213) regards it as the first example of dark matter in the history of cosmology. And the Latin church Credo speaks of all things seen and unseen, visibilium et invisibilium. Copernicus and Newton appear to have had no use for the invisibilium, but the idea was at least available for discussion. The decades from 1785 to 1845 brought forth a pair of ideas (black holes and dark stars) and a pair of observations (white dwarfs and Neptune) that are direct ancestors of some recent dark matter candidates. The year 1800 also saw the first deliberate search for something that might previously have just been too faint to be noticed. Only one of the seven items alluded to so far (entities with escape velocity larger than the speed of light) meets the present definition of dark matter: stuff that neither emits nor absorbs its fair share of radiation at any wavelength, but can interact with it gravitationally and perhaps weakly.

1.1 Paleomatter

The first idea was "unenlightened stars," which might orbit and eclipse visible stars, producing periodic variability, first seen in Algol. Published by Edward Pigott (1805), the idea had been discussed between him and John Goodricke back around 1786 (Hoskin 1997, p 202). The second idea now carries the name "black holes," a phrase which first appears in print in Science News Letters for 19 January, 1964 (p. 39), but which gained popularity only later when used by John A. Wheeler in talks and papers (1968). The concept was phrased by John Michell (1784) as "all light emitted by such a body would be made to return to it by its own proper gravity," and by Pierre Simon de Laplace (1796) as "il est donc possible que les plus grands corps lumineaux de l'universe soient par cela meme invisible."

The pair of observations occupied a number of years, but the analyses and announcements were nearly simultaneous (at least in our rest frame) in the 1840's. F.W. Bessel (1845a, b) concluded that the paths of Sirius and Procyon across the sky were not straight, but oscillatory, so that each must have an invisible companion, and U.J.J. Leverrier (1847, 1848) decided that irregularities in the motion of Uranus around the sun implied the existence of a more distant planet. J.C. Adams (1847) had reached a similar conclusion in the same time frame as Leverier, but it was the latter's work that led to the 1846 discovery of Neptune by Galle at Berlin. The companions of Sirius and Procyon were discovered in 1862 (accidentally for Sirius by Alvan Clark) and in 1896 (after deliberate looking around Procyon by Schaeberle, 1896).

In the meantime, F.X von Zach (Hoskin p. 188-190) attempted in 1800 to organize a team of 24 "celestial police" to scour the skies for the planet that, according to Bode's law, must come between Mars and Jupiter. They were scooped by Piazzi's accidental discovery of Ceres. He had not yet been told that he was supposed to be one of the policemen, though not for that zone. This is not untypical of much of the history of dark matter and, indeed, astronomy in general.

1.2 Some Characteristics of History of Science

First, it can never be complete. So far we have touched on 7 possible entities and concepts – counter-earth, the invisibilium, unenlightened stars, black holes, white dwarfs (like Sirius B), gas giants (like Neptune), and small chunks of solid baryons (like Ceres and her smaller sisters found later), without having yet reached the 20th century. They do not seem to have been brought together in any of many previous discussions of dark matter history, though all but the first have been suggested as DM candidates for recent decades. Kragh (2007) notes Philolaus and black holes; Einasto (2001) begins in 1915, and van den Bergh (2001) in 1933. My own previous ventures into this territory (Trimble 1987, 1988, 1995, 2005) also pick out only subsets.

Second, comparing the isotemporal articles by Einasto and by van den Bergh leads to the conclusions that claiming anything as "the first" is very dangerous, and that history can look very different to folks on opposite sides of borders (the Oder-Neiss line in this case).

In addition, the professional historians themselves are not in full agreement about how one ought look at things – in any case, not just "who got it right first" (which is called Whiggish) or "who persuade the community" (populist), but perhaps more nearly, "why did X think Y, who influenced him, whom did he influence, and how?" The approved perspective changes with time. I think the concept of a paradigm shift (Thomas Kuhn in 1962) had already gone out of fashion by the time Scott Tremaine (1987) described the recognition of dark matter in those terms. He put the beginning of the shift in 1974 and predicted that things should be sorted out by 1994 (perhaps just barely true, though I date the current standard model to 1997 at the Kyoto General Assembly of the International Astronomical Union).

Fourth and finally, progress is hardly ever monotonic. How did two such sensible ideas as eclipses by dark bodies and $2GM=Rc^2$ ever get dropped from the astronomer's tool kit? In the case of the stars, observations of additional ones (especially Delta Cephei) revealed variables with imperfect periodicity, non-symmetric light curves, and other irregularities that Keplerian orbits could not match, leaving the older idea of spotted, rotating stars (like the sun only more so) master of the terrain until the late 19th century. Binaries were then the best buy model for all variables until analysis of Delta Cephei and others showed that one star would have to be inside the other, leading to pulsation as a third mechanism. All three, of course, actually happen, occasionally in a single system.

As for the bodies that held back their own light, the problem was that $2GM/Rc^2 = 1$ had been calculated with a particle theory of light, Newtonian mechanics, and the assumption that the process would manifest itself as light being slowed down as it left massive sources. The wave theory of light took over in the early 19th century, leaving no obvious way to calculate the phenomenon for bodies with less extreme M/R. Observationally, the slowing of light would have shown up as color (or at least Doppler) shifts and as loss of synchronism of orbits of the two stars in visual binaries. Nothing of the sort was seen. Thus the whole realm of interaction of light with gravitational fields had to be rethought in the light of special and general relativity, beginning with Einstein himself (deflection of light; gravitational redshift), Chwolson (gravitational lensing), Schwarzschild, and all the rest. That the classical and relativistic expressions for the Schwarzschild radius and (to first order) gravitational redshift are the same counts as one of the minor mercies of 20th century astrophysics. Light bending, which had also been calculated in the particle/Newtonian era, is a factor of two larger in the relativistic case. This comes into the evaluation of alternatives to dark matter in their ability to describe the velocity distributions, X-ray temperatures, and gravitational lensing for clusters of galaxies.

Notice that, in contrast to the theoretical cases, the observed entities have never been out of the astronomical inventory. As for the "celestial police," many astronomers since have tried to carry out major programs by persuading their colleagues to help out (the Carte du Ciel in the 1880s; Zwicky's supernova search programs in the 1960s). There must be a better analogy than cat-herding, but it is really only in the last decade or two that large teams have come together in large survey or monitoring programs using many facilities, some purpose-built. The driver has been the need for a large constituency to pry loose large numbers of euros, dollars, and all. Recent dark matter searches have begun to approach this condition.

2. When Stellar Statistics was the Astronomical Forefront

Each era has had its handful of astronomical questions generally regarded as important. Today's seem largely to be in the territory of origins and formation – of the universe, of galaxies, of stars and planets - with dark matter (and dark energy) part of at least the first two. New observations and interpretations in each epoch will generally make sense only within the framework of those important questions. Once upon a time, from the Greeks to Tycho and Kepler, to Newton and Halley and beyond, it was obtaining accurate observations of the moon, sun, planets and their moons, and comets and fitting them with a theory that would permit prediction of eclipses, transits, conjunctions, and other mutual events and returns. Through most of this period, the "fixed stars", apart from an occasional nova stella or variable, were merely the pattern against which important motions could be measured, though Newton was of the opinion that Divine Intervention might be necessary to keep them fixed for long periods (Kragh 2007, p. 73 on 1892-97 correspondence with Nicholas Bentley).

2.1 The Stars in Motion

But by 1700, the stars were about to break loose from their moorings, in four steps. These, in chronological order of discovery (broadly interpreted!) are called proper motions, binary stars, parallax, and radial velocities. First, Edmund Halley (1718) announced that three of the brightest stars had coordinates different (above and beyond precession) from those reported by the ancients. James Bradley's 1729 recognition of aberration of starlight is to us a triumph of precision measurement which required the earth to orbit the sun rather than conversely. But it was then also a major source of noise in efforts to determine accurate stellar positions and motions. The names associated with the gradual accumulation of meaningful proper motions (Hoskin 1997, p. 202-209) were Bessel, Tobias Mayer (who pointed out in 1760 that it was relative motion of the star and sun being seen), William Herschel, F.W.A. Argelander, and Thomas Galloway. All attempted to discern the motion of the sun relative to all the other stars; their answers were largely concordant (including that of Galloway, who used an independent southern sample) and in agreement with the modern solar apex, toward Hercules.

Bradley had, of course, been looking for parallax, and so was Herschel when he (1803) recognized that a few close pairs of stars in the sky were moving around each other. That there were too many close pairs and clusters to be chance superpositions had been noticed and published by John Michell (1767) but largely forgotten. That the orbits showed Newtonian gravitation beyond the solar system became certain with the first eccentric pair (Xi UMa, charted by Felix Savary in 1827).

The long-sought parallax appeared in three sets of measurements of three different stars by two famous astronomers (Bessel yet again and Wilhelm Struve) and one rather obscure one (Thomas Henderson who had the advantage of working from the Cape of Good Hope and looking at Alpha Centauri) in 1835-38. The values found, all less than one arc second, confirmed the very large distances (10^6 AU or parsecs and more, though neither unit yet had that name) implied by apparent stellar brightnesses and the idea that stars were suns. Of course observing the sun and other stars at the same time presented certain difficulties! Huygens tried putting a screen with a pin-prick hole between himself and the sun.

But the hole was too big, and he ended up putting Sirius at only 27,664 AU. James Gregory in 1688 suggested using a planet as an intermediary and got 83, 190 AU (Hoskin p. 211), and Newton did still better but did not publish until 1728. The modern value of course exceeds 200,000 AU, but apparent brightnesses and then parallaxes in any case made clear that the "universe of stars" was very large indeed compared to the solar system. Though the inventory of reliable parallaxes also grew very slowly, they enabled the conversion of proper motions to linear speeds and the use of visual binary orbits to measure a few stellar masses, which indeed proved comparable with that of the sun.

Radial velocities required the development of spectroscopy and so came last of the four. Huygens tried but failed (again), and the first handful of meaningful numbers-meaning errors of a few km/sec for values of 10-30 km/sec - came from H.C. Vogel and Julius Scheiner at Postdam and James Keeler at Lick around 1890. The first spectroscopic binary orbits followed soon and provided a few more stellar masses.

With the stars now in motion, it became possible to ask about the statistics of numbers vs. luminosity, distance, mass, proper motion, and radial velocity; to calculate the amount of mass necessary to account for the motions; and to ask whether the stars themselves added up to the necessary mass

2.2 Mapping the Milky Way

It is in this context that astronomers first found some hint of dark matter in the Galaxy (or universe) in the early 20^{th} century. In the same time frame, determinations of the solar motion relative to, first, more distant stars, then globular clusters, and finally external galaxies led up to the velocity-distance relation we call Hubble's law.

None of the pre-1918 data strongly contradicted the widely-held opinion, going back at least to Herschel, that the solar system was near the center of a thick disk of stars a few thousand light years (or parsecs, to within a factor three!) in diameter. Many of the best-known astronomers of the period worked on problems of stellar statistics, including Kapteyn, Jeans, Eddington, Karl Schwarzschild, Strömberg, and later Oort. They invented descriptors like star streams, velocity ellipsoids, and asymmetric stellar motions. All would eventually be subsumed in the idea of galactic rotation (of the disk, but not the halo) and velocity dispersions around the average rotation in radial, angular, and perpendicular directions, R, θ, and Z.

An interesting snapshot of the community struggling with these various concepts, after the recognition by Shapley of the off-center position of the sun and by Hubble of the existence of other galaxies, but before the discovery of rotation, is found in Russell, Dugan, and Stewart (1926, Vol II), which was the primary astronomy text for a generation of our English-speaking predecessors. Russell et al. reproduce an argument due to Kelvin, which says that the volume occupied by a uniform density of stars like that near us cannot be either very small or arbitrarily large, or stars would either leave, in the small case, or have much larger speeds than the largest seen (about 350 km/sec then) in the large case. Conversely, the largest speeds plus a Galaxy size of Herschelian or Kapteynian dimension require an average density of about one low mass star (the commonest sort) per cubic parsec. Russell et al. state specifically that the argument applied to the density of "all matter, whether luminous or dark."

This was the context in which Ernst Opik (1915) set out to find the density of matter near the Galactic plane, using the speeds of stars and their distances from it. He concluded that the required mass might well all be in the stars themselves, assuming a mass to light ratio of 2.63 (in solar units) implied by counts of stars of various spectral types. He combined his result with the apparent surface brightness of the central part of the Andromeda Nebula, velocity information from Slipher's (1914) spectrum, and Newton's laws to deduce a distance of 440 kpc for M31 (Opik 1922), about half the current value and obviously well outside the Milky Way.

Notice that Opik was essentially right both about the minor importance of local dark matter and about the existence of external galaxies at a time when the majority of the astronomical community would have voted with writer and popularizer Agnes Mary Clerke, whose pair of books in 1896 and 1903 (Hockey et al. 2007, entry for A.M. Clerke) denied with great vehemence the existence of other observable galaxies but included a whole chapter called "Dark Stars." By this she meant ones that had faded following exhaustion of their (unknown) energy supply. We call them white dwarfs, and they have been dark matter candidates in the past.

The story of stellar energy production is an almost completely separate one from the dynamical issues addressed here, but Simon Newcomb (1906), first president of the American Astronomical Society, described the source of solar and stellar energy as the most important unsolved astronomical problem of his time, though his own work was largely in positional astronomy. He also questioned the central position generally assigned to us in the Galaxy (by analogy with Ptolemaic cosmology) and asked himself whether the perihelion advance of Mercury might perhaps be attributable to a small deviation from Newtonian gravity.

3. Dark Matter Between the Wars: Some Pairs of Papers

Let us deal first with the divisive question of who coined the phrase dark matter. Zwicky (1933), whom you will meet properly in Sect. 3.2, writing in German, spoke of "dunkle Materie" and is often given credit, but we will see "dark matter" in English appearing a decade earlier. The name has taken a very long time to become fully standard. In 1977, Ivan King, Martin Rees, and James Gunn (separately, in Tinsley & Larson 1977) all wrote "missing mass" or "unseen mass. J.P. Ostriker enunciated at that meeting, before, and after it, that it is the light that is missing, not the mass. King countered that the mass was in any case "missing from our understanding" and stuck by the old phrase. Dark matter was, however, the winner over the next few years (Trimble

1987). There have also been holdouts against the very notion, some of whom appear in later sections. If you are young enough that 1977-87 is history for you, rather than current events, then you may even need reminding that the two wars were those of 1914-18 and 1939-45. And if someone should try to sell you a document said to have been written before 1940 which speaks of World War I, it is a forgery.

3.1 The Galactic Disk: Kapteyn and Jeans

In the years after 1915, the number of accurate radial velocities, stellar distances, and masses accumulated rapidly. Some of the work was driven by J.C. Kapteyn's effort to map out the entire Milky Way, using his Selected Areas, and his (Kapteyn 1922) estimate of the mass in the Galactic disk is a fairly small part of that program. The 1922 paper was written only about a year before his death, and his decision to assume no interstellar absorption or scattering of starlight was perhaps driven by the feeling that he must finish the paper soon or not at all. Jeans (1922) also used velocities and distances perpendicular to the plane to calculate the mass density in it, without at that time attempting to map the entire Galaxy.

Their calculations of the local mass density are not invalidated by their assumption that the solar system was very near the center of the Milky Way, since only stars within about a kiloparsec were used. Their results (0.143 and 0.099 $M_\odot / pc^3$) roughly bracket more recent values, as do their conclusions, based on comparing the mass per luminous star implied by their numbers with the average mass of known binary systems (1.6 $M_\odot$ at that time).

Jeans opined that "there must be about three dark stars in the universe for every bright star," while Kapteyn, noting that "we have therefore the means of estimating the mass of dark matter in the universe," concluded that "as matters stand at present, it appears at once that this mass cannot be excessive." Their universe is, of course, our Galaxy, and was not capitalized.

The Jeans and Kapteyn conclusions are, more or less, phrased respectively as a detection of dark matter and as an upper limit. The two came gradually together, once faint stars, stellar remnants, and gas were promoted from dark to enlightened and measurements of stellar distances outside the plane became more numerous and accurate. We are, incidentally, less than 20 pc from the median plane.

The current status of dark matter in galactic disks is that they will have a slightly larger density than the adjacent halo because of dark matter brought in, along with baryons, in mergers and satellite captures (Read et al. 2008). Two short-lived alternatives were the possibility of very dense, compact clouds of molecular hydrogen in outer disks (ruled out by absence of absorption in them, Clarke et al. 2003) and outer light that had been missed (Valentijn 1990), eventually ruled out by deeper infrared imaging.

3.2 The Two Most-Cited Early Papers: Oort and Zwicky

More data for more stars continued to appear, and what Oort (1932) brought to the table was a much more sophisticated way of analyzing the data, involving Poisson equations and so forth. He had by then also provided from stellar motions strong evidence for galactic rotation (Oort 1927), which had been proposed the year before by Lindblad (1926). The most accessible discussion of the methods is probably Oort (1965), in which he describes the goal of stellar dynamics as finding relations between the density and velocity distributions of stars. The 1932 result for local mass density was 0.08 to 0.11 $M_\odot$ / pc^3, so that one solar luminosity corresponded to 1.8 solar masses. Stars brighter than M_V = 13.5 contributed 0.038 $M_\odot$ / pc^3. Oort's opinion then was that stars between M_V = 13.5 and 18.5 would account for most of the rest. This was based heavily on apparent (and erroneous) gravitational redshifts near 240 km/sec for the white dwarfs Procyon B and van Maanen 2, which (using R.H. Fowler's non-relativistic equations for degenerate matter) would have required them to have masses much larger than 1 $M_\odot$.

Oort's 1965 number was 0.148 $M_\odot$ / pc^3, of which about 40% had to be attributed to stars or gas of unknown type. (HI had been inventoried by that time, but not molecular hydrogen.) This number is still often called the Oort limit, though by now we see that it should really be the Kapteyn-Jeans limit or even the Opik limit.

In addition to the local mass density, one would like to know M and M/L for a cylinder extending far out perpendicular to the disk, so that the numbers can be compared with those for other galaxies as well as telling us something about the nature of dark matter in our own. Oort (1965) pointed out that, if the unknown 40% were distributed like halo stars, then M/L in the cylinder would be as large as 13.4 (but only 2-6 for the 40% all in a thin plane distributed like old disk stars). He regarded the larger number as almost impossible, though it was comparable with what de Vaucouleurs (1958) had found for a similar cylinder about 10 kpc from the center of M31. We would now say that these large numbers made sense if the cylinders extended to 100 kpc, through the galactic halo, but they did not, because there are no tracer stars of the disk population more than a couple of kpc out of the plane.

Before moving on to the other, Zwicky, paper, recall that, for anything dominated by gravitation, a reasonable mass estimate is $M = V^2 R/G$ for properly selected velocity, V, and length scale, R. V is mercifully distance-independent, but R (hence M) is, of course, linear in distance and, therefore, in the reciprocal of the Hubble constant H, for things outside the Local Group. Luminosity corresponding to measured apparent brightness goes as R^2, so that M/L values for extragalactic entities are linear in H. Between the first 1929 Hubble paper and the Rome 1952 IAU General Assembly, nearly everyone took H to be close to 500 km/sec/Mpc (Trimble 1996 mentions a few early exceptions). Inverted, this is the same arithmetic Opik (1922) used to estimate the distance to M31. Hubble's (1934) estimate of $10^9\ M_\odot$ for a typical galaxy assumed M/L like the solar neighborhood (i.e. no dark matter) and included only central bright regions.

No recent astronomer seems to have doubted the existence of galaxies outside ours, but reservations about clusters persisted. Hubble thought they were rare, while Shapley thought truly isolated galaxies were rare. Partly they differed in definition: Hubble meant things like Coma and Virgo, while Shapley included the Local Group, which from 100 Mpc would look like a binary of the Milky Way and M31. And, thought Shapley, Hubble had access to excessively large telescopes, whose small fields of view made it hard to see the big picture. Indeed "cluster denial" has persisted almost to the present. For instance, Fasenko (1985) and Zabrenowski (1986) and other papers by the same authors attribute the patchy distribution of galaxies on the sky to differential absorption in the Milky Way, in which case measured velocity dispersions are merely random fluctuations around the Hubble flow.

The second now highly cited paper is that of Zwicky (1933), who realized that for a cluster of anything the relevant V is the velocity dispersion and turned the 100" telescope at Mt. Wilson toward the Coma cluster. He measured eight radial velocities and found a dispersion of about 1000 km/sec, larger than he had expected, and leading to his conclusion "dass dunkle Materie in sehr goesserere Dichte vorhanden is als leuchtenden Materie." That is, a thousand "Hubble" galaxies of $10^9\ M_\odot$ each would add up to only $10^{12}\ M_\odot$, and he had found something more like $10^{14}\ M_\odot$. This number is not called the Zwicky limit even now, though he is generally now cited as a dark matter pioneer. Zwicky supposed that dwarf galaxies and gas must make up the missing 99%, later adding intergalactic stars and pygmy galaxies to his inventory. He looked for evidence of these as excess sky brightness between the known Coma galaxies (Zwicky 1951 and 1956) and thought he had seen an appropriate number of photons. Intracluster light and indeed long-lived entities like globular clusters and planetary nebulae have been reported many times since, but they do not add up to more mass than is in the recognizable galaxies (Da Rocah and Mendes de Oliveira 2005).

A few years later, Sinclair Smith (1936) used a spectrograph of his own design to make the Mt.Wilson 60" nearly as powerful as the 100" for some purposes. He aimed the combination toward the Virgo cluster and, from 32 radial velocities, found a total mass of $10^{14}\ M_\odot$ (for a distance of 2 Mpc, thus about $9 \times 10^{14}\ M_\odot$ at a modern distance). He noted that the “implied mass per nebula is 2 x 10^{11} $M_\odot$...far larger than Hubble’s value of 10^9 $M_\odot$ for the mass of an average nebula.” He went on to say that “it is possible that both figures are correct and that the difference represents a great mass of internebular material within the cluster.” Indeed it does.

3.3 Two Intermediate Scales: Holmberg and Babcock

Holmberg’s (1937) thesis primarily addressed the demonstration that there must be bound pairs of galaxies, because there were too many close pairs on the sky to be chance superpositions, very much like Michell’s 1767 argument for binary stars, but Holmberg did a better job of the statistics. As a by-product of the measured pair separations on the sky and in radial velocity, he ended up with an estimate of average mass per galaxy or M/L for binaries His average mass was 10^{11} $M_\odot$ (perhaps 5 x 10^{11} $M_\odot$ on a modern distance scale). Remarkably, given the large errors everywhere, his number fit comfortably between Hubble’s single galaxies and Zwicky’s Coma cluster. He cited Zwicky (1933) and Smith (1936), as well as Hubble and Shapley, and stated that his value was a satisfactory one “in very good agreement with the values earlier derived.” He remained an outstanding extragalactic astronomer throughout his career.

Horace Babcock (1939), who measured the first extended rotation curve for a spiral galaxy (M31 naturally), left the field and became primarily known for contributions to solar physics, especially the theory of the 22 year activity cycle. He explained why in a handwritten letter responding to my inquiry about why no-one (including himself) had followed up on his work for such a long time. His rotation curve was still rising at the last measured point and (scaling to a distance of 785 kpc) implied M = 3 x 10^{11} $M_\odot$ and M/L = 17 out to 18 kpc. His paper remarked upon the difference from the Milky Way as then understood and did not cite Zwicky, Smith, or Holmberg. A public talk on his results (at the 1939 dedication of McDonald Observatory) was heavily criticized by other meeting participants, and he was instructed to publish his thesis as a Lick Observatory Bulletin rather than in the Astrophysical Journal. He found neither of these events encouraging, though curiously his contemporary, Daniel M. Popper, also at Lick was instructed to publish in ApJ, when he would have preferred the LOB series! Babcock’s last data point, 380 km/sec 100’ from the center of M31 came from a multi-night exposure and was probably erroneous, though the resulting mass and M/L now sound right on. Vera Rubin (personal communication and talks at meetings) has wondered whether he might perhaps have had doubts about the numbers even at that time.

The second rotation curve ever measured was that of M33 (Mayall and Aller 1940). M33 is now known to be tidally distorted and truncated, so that the small mass and M/L they found were inevitable. It is hard to say in retrospect whether this was good or bad luck.

3.4 Inter-war Summary

A widely-read astronomer of trusting spirit could, in 1939, have plotted a little graph, like Figure 1, showing a monotonic, nearly linear increase of M/L with length scale (at least in log-log coordinates!). This was not done at the time and perhaps would not have suggested to 1939 eyes dark matter more widely distributed than the luminous. There was one version in tabular form due to Schwarzschild (1954) which he did not interpret as an M/L gradient, and such plots did not appear for another 20 years. Indeed none of the data were much discussed, and van den Bergh (2001) presents in some detail the extent to which the Zwicky (1933) and Babcock (1939) papers were ignored. Zwicky's (1937) suggestion that gravitational lensing by galaxies would firmly settle the issue of their masses belongs chronologically to this period but intellectually to the modern, dark-matter dominated era.

4. INTERLUDE: 1939-61

Two fairly obvious things kept this generation from being a golden one of dark matter research. First, there was a war going on and then a good deal of essential post-war reconstruction, especially in Europe (from Angelsey to Kiev). Second it was a golden age of stellar structure and evolution research, because a basic understanding of nuclear reactions as the energy source for the sun and stars had been achieved just as WWII broke out and because computational power was rapidly expanding during and after the war. Thus, many outstanding astronomers focused on stars. As for our pioneers, Jeans (1877-1946) and Kapteyn (1851-1922) had died full of years and honors and Sinclair Smith (1899-1938) tragically early of cancer. Zwicky (1898-1974) and Oort (1900-1992) continued to advocate dark matter and many other things up essentially until the times of their deaths.

Nevertheless, data suggestive of large M/L ratios in various contexts gradually accumulated. The contexts included the Milky Way, other individual galaxies, pairs (including the Local Group), rich clusters, and the universe as a whole. There were also papers that found mass-to-light ratios fully accounted for by normal stars and gas, most notably a series by E.M. and G.R. Burbidge (1961 and references therein) on rotation of spiral galaxies, in which, however, they repeatedly emphasized that they could say nothing about faint outer regions where the photographic plate detectors of the time did not permit recording either emission or absorption line spectra. In case you had ever wondered, Slipher's spectrum of M31 showed absorption features from the stars most clearly and Babcock's the emission lines of HII regions.

The other distractor was a pair of papers on binary galaxies by Page (1952, 1956). He differed from Holmberg in requiring smaller separations on the sky and in radial velocity before declaring a pair to be physically bound. Smaller average mass per pair was an inevitable result.

Items are arranged by length scale rather than by year in the rest of this section.

4.1 The Milky Way

Calculated masses for our Galaxy interior to the solar circle depend on the adopted values of local rotation speed (V_c) and galactocentric distance (R_o), with detailed modeling of flat vs. spheroidal distributions less important. Thus this mass has shrunk from somewhat larger 10^{11} $M_\odot$ to somewhat smaller over the past few decades, as best estimates of V_c have declined from about 250 to about 220 km/sec and of R_o from 10 to 8-8.5 kpc (Ghez et al. 2008). But it is the mass outside R_o that is important for dark matter, and this is much less certain. Something near 1 x 10^{11} M_o for the inner Galaxy would have been fine with Oort (1932) and also Kuzmin (1952) and Schmidt (1956) among the early modelers.

The population of globular clusters extends considerably beyond the solar circle, and the dispersion of their velocities provides a handle on the larger scale mass. Knowing only the radial component is a challenge, but Kunth (1952) and Lohman (1956) each found two or three times as much mass outside R_o as inside, and more elaborate analyses confirmed their conclusion (Table 1 of Trimble 1987), though with results spread across 0.3 – 30 x 10^{11} $M_\odot$. More recently, proper motions for some globular clusters have been measured, and the inventory of dwarf galaxy companions has expanded to the point where they can be used the same way. The conclusion that outer (halo) mass considerably exceeds inner (disk + bulge) mass holds.

4.1 Other Individual Galaxies

The third published rotation curve was Oort's (1940) for the S0 galaxy NGC 3115. He stated that "the distribution of mass in this object appears to bear almost no resemblance to that of light" (a view which the Schwarzschild 1954 paper was partly designed to refute). Van den Bergh (2001) attempts to understand Schwarzschild's view, but the explanation bears some of the signs of a cracking paradigm – large error bars, data that should be left out of the sample, and emphasizing data that agree with the pre-determined conclusion, like the rotation curve of M33. Schwarzschild did accept that the visible parts of elliptical galaxies had larger M/L ratios than the corresponding parts of spirals. He attributed the difference to a large population of white dwarfs in the ellipticals. Oort's candidate for NGC 3115 was faint M dwarfs.

The advent of radio astronomy would eventually settle the issue for rotating galaxies, beginning with van de Hulst's (1957) confirmation of a global M/L near 20, vs. 2 in the nucleus (Lalleman 1960) of M31. There do, of course, also exist some rotating galaxies with rotation curves that turn over at large radii, (Cicaire et al. 2008) and the case for dark matter in ellipticals was historically more difficult to make because one needed to use absorption lines that are the sums of many stars on sight lines through various parts of the galaxy, until radial velocities for their globular clusters, x-ray temperatures, and lensing data (Pooley et al. 2008), X-ray temperatures became available.

4.3 Binary Galaxies and Small Groups

Page (1960) reconsidered his sample, corrected some numerical errors, and endorsed Holmberg's (1937) larger M/L values. So, not so surprisingly, did van den Bergh (1960) and Holmberg (1954) himself, using larger samples and a less stringent criterion for which pairs to count as bound. At least six papers (cited in Trimble 1995 Sect. 4) analyzed small groups and found mass-to-light ratios up to about 100.

The most innovative discussion was surely that of Kahn and Woltjer (1959), who brought together the ideas that our Galaxy and Andromeda are approaching each other (known to Slipher, 1914) and that the universe is expanding (if not quite known to Hubble in 1929 then to others soon after). The actual approach speed is about 100 km/sec, part of the observed 300 km/sec being due to Milky Way rotation; and the Universe of 1959 was about 10 Gyr Old. In that time, the gravitating masses of the two galaxies must have been enough to stop their motion apart from each other and turn them around to start coming back together again. Not knowing the relative transverse speed or whether the upcoming close passage will be the first were obvious sources of uncertainty, but could not force the total mass far away from 10 times the numbers for the inner 10 kpc of each.

The authors supposed that the other 90% ought to be gas in some temperature/density regime not (yet) detectable. Oddly, in the proceedings of IAU Symposium 35, Woltjer is quoted as saying that this 10^{12} $M_\odot$ of gas was "a number that had crept into the literature and now seems to be creeping out." Much of the gas, was, as it were, forced out by advances in radio and X-ray astronomy, but there is arguably still some supply at about 10^5 K that continues to fall in as high velocity clouds and to impose ultraviolet absorption lines of multiply ionized oxygen (etc.) on the spectra of more distant sources. This belongs to the "missing baryon problem" (Richter et al. 2006).

4.4 Rich Clusters

Between 1950 and 1960, the Virgo and Coma clusters were re-examined at least three times each and the clusters in Hercules and Canum Venaticorum at least once each (references in Trimble, 1995, Sect. 4). As a rule, there were 10's of radial (one-dimensional) velocities out of a thousand galaxies, no transverse velocity information, and the best value of Hubble's constant was declining rapidly from 500 to perhaps 125 km/sec/Mpc (Trimble, 1996). Nevertheless there was striking general agreement that cluster mass-to-light ratios ranged upwards from 100 to perhaps as much as 1000. Recall that, for a cosmic luminosity density of $10^8 L_\odot$ / Mpc^3 (another of those numbers that goes back to Oort and hasn't changed much), M/L a bit larger than 1000 takes us to the critical, closure density for a standard relativistic universe.

4.5 The Universe as a Whole

Einstein's equations permit a very wide range of values for the Hubble constant, H, the density and pressure of stuff, ρ, and the cosmological constant. Early cosmologists (Tolman 1934, for instance) considered many possibilities including some with no epoch of very high density and oscillating models that cannot apply to our universe.

In contrast, the Steady State universe of Bondi and Gold (1948) and of Hoyle (1948) requires precisely the critical density at all times, presumably in hydrogen gas available to form new galaxies, at a temperature greater than 10^5 K (Hoyle 1959), which at the time could not have been ruled out observationally. This requirement for a very large amount of undetected matter was one (though not the only) source of violent objections to the model (Kragh 1996). Other alternatives to Einsteinian cosmology, e.g. those of Dirac and Milne, also required a critical density (Kragh 2007, Sect. 3.4). Dirac held by such a model through most of his life, though the version he spoke about at a conference in the 1970s at a planetarium dedication would have given us a sky glowing with H-alpha emission that you could almost read by. This was also the meeting where, at luncheon, he asked his neighbor at the table, "Do you want my ice? Ice causes flu."

Despite the wide range of possible general relativistic universes, the choices $\Lambda = 0$ and $q_o = ½$ (critical density) were very popular in mainstream texts and articles. There seem to have been three reasons for the choice: (1) those conditions could not be ruled out (unless you worried about ages of things), (2) k=0 (flat space) and $\Lambda = 0$ together have an esthetically pleasing appearance, and (3) the calculations needed for analyzing data (apparent magnitudes and angular diameters vs. redshift, for instance) are thereby much simplified. Non-zero k's and Λ's, however, also appear scattered through the mainstream literature, motivated by considerations of the formation of helium in the early universe, ages of things vs. various values of the Hubble constant, and the difficulties of forming galaxies and clusters in an expanding universe.

5. 1961-74: From a pair of conferences to a pair of short papers

At least half a dozen items relevant to dark matter appeared during this period, including non-Newtonian gravity; X-ray emission from rich clusters and the X-ray background; a revival of gravitational lensing; discovery of the 3K background and implications for galaxy formation from its extreme isotropy; two new dark matter candidates (neutrinos of non-zero rest mass and primordial black holes); extended rotation curves; and Big Bang nucleosynthesis. Subsequent events that have changed the relationships of these items to the primary issue of existence and nature of dark matter are mentioned sporadically.

5.1 The 1961 Conferences

The International Astronomical Union General Assembly of 1961 was the first held in the USA and the first with associated symposia before and after. IAU Symposium 15 (McVittie 1961, McV below) ranged over the general territory of extragalactic astronomy and includes a brief remark by Zwicky. An additional associated meeting was specifically targeted at “The instability of clusters of galaxies” (Neyman, Page, & Scott 1961, NPS below). Zwicky was not a speaker, was referenced in only one meeting presentation, and organized an unofficial evening session to present his own ideas, but then spoke largely in opposition to the idea of superclusters of galaxies rather than about dark matter, according to Vera Rubin, who was there.

A wide range of views was expressed at the two events, most of which had a few adherents up to the 1980s (Trimble 1995, Sect. 5), but dark matter was, of course, the winner. Holmberg (in NPS) hoped that the problem would just go away and suggested that observational errors, foreground/background interlopers, and substructure in clusters could account for large velocity dispersions. Lemaitre (NPS, p. 603) proposed that rich clusters might continuously exchange galaxies with the field, so that the configurations were long-lived, but the individual members need not be gravitationally bound together.

Ambartsumian (NPS p. 536 and references therein) was firmly committed to the idea that clusters of galaxies were expanding out of some denser, more compact configuration, in somewhat the same way that some star clusters are clearly unbound and expanding. Supporters in 1961 included Vorontsov-Velyaminov (NPS p. 551), Kalloghjian (NPS p. 554), Markarian (NPS, p. 555), de Vaucouleurs (NPS p. 629), and the Burbidges (NPS p. 541).

Even then, the analogy with star clusters could have been only an approximate one, because we see other star clusters clearly bound by the stars themselves, and very young ones are forming from gas clouds that were initially contracting but dissipate when kinetic energy is added by winds and supernovae of the most massive stars. And we see no 10^{17} $M_{\odot}$ gas clouds that could be the immediate precursors of unbound clusters to form in the future. In fairness to Ambartsumian, he never supposed we did. His image was more nearly akin to that of Jeans, who once described the spiral nebulae as places where material was pouring into our universe from elsewhere. In fact, Ambartsumian's image of stellar evolution was a similar one, where compact configurations came first, and dense pre-stellar matter expanded to make the stars (with planetary nebulae as an early phase rather than a late one in stars' lives). A crucial issue is that rich cluster galaxies, especially in cores, are mostly ellipticals, and field galaxies more often of later types, a point emphasized by van den Bergh (1961, NPS p. 566). He also briefly considered a possible role for non-gravitational forces (also the view of de Vaucouleurs 1960), but rejected it.

There was, however, also a "stability of clusters" coterie at the meetings. Zwicky (McV p. 347) and Baum (McV p. 255) thought there might be enough intracluster starlight to bind the clusters with luminous material. Minkowski (NPS p. 558) and Limber (NPS, p. 572) were also on the "bound side." Most committed was Abell (NPS. P. 607). He had taken a large fraction of the Palomar Observatory Sky Survey plates, examined them in detail for clusters and other structures, and compiled a catalog of clusters (still in use). The process had persuaded him that not only clusters but "second order clustering of galaxies and interactions between clusters of galaxies" existed, on length scales of 50 (100/H) Mpc and with masses of 10^{17} (100/H) $M_{\odot}$. Zwicky was a firm opponent of superclustering, but some of his clusters (in an independent catalog also derived from POSS plates) had sizes of 40 (100/H) Mpc and enough substructure that an innocent bystander might well have said they displayed second order clustering.

Van Albada (NPS p. 590) and von Hoerner (NPS p. 580), both better known for other things, attempted to determine the extent to which the Virial theorem could be applied to rich clusters when only a handful to a score of redshifts had been measured. Using analytical and numerical methods respectively, they concluded that Virial masses were likely to be wrong, but not by factors of more than two or three after significant relaxation of the clusters. Von Hoerner's calculation was plausibly the first N-body simulation in astrophysics. It had N=16.

5.2 X-ray Considerations

These give us numbers for both gas and total mass of rich clusters and limits on cosmic diffuse hot gas. The first extra-solar-system X-ray discoveries were Sco X-1 and a very roughly uniform background (Giacconi et al. 1962). The standard pioneering discussion of possible emission mechanisms is that of Felten & Morrison (1966), who considered among other possibilities thermal brehmsstrahlung of ionized hydrogen, in an amount that could, just about, close the universe (for H around 50) without violating the Gunn-Peterson (1965) limit on neutral hydrogen in ionization equilibrium with it. Other calculations established that even if the entire background were due to hot, ionized hydrogen, the universe could not be closed that way (Hoyle 1963, Gould & Burbidge 1963; Field & Henry 1964). Ruling out a significant contribution was, however, a slow process, driven by the gradual increase in the fraction of the observed background attributable to resolved sources and the angular fluctuations of the rest (Gursky & Schwartz, 1977, reporting when the issue was just about resolved).

In the rocket X-ray era, there were tentative detections of one or two clusters of galaxies as X-ray sources. The Uhuru satellite reported something like 40 in 1971-74. Again the theorists considered all the usual radiation mechanisms, but in this case detection of an iron emission line settled the question in favor of thermal bremsstrahlung. Early estimates of the amount of gas required in some cases came close to the total needed to bind the clusters, but it soon became clear that the gas mass is about 10% of the total, comparable with the mass in galaxies (again settled by the time of the Gursky & Schwartz 1977 review). But, of course, the hot gas must be contained through the life of the cluster, and the mass required to do that proved comparable with total cluster masses found from galaxy velocity dispersions and the Virial theorem, providing an independent demonstration of dark matter in rich clusters. An easy, approximate way to think of it is that kT/m_e is essentially the appropriate velocity for an $M=V^2R/G$ mass, where R is some measure of cluster size. Whether the masses found in the two ways are the same, or should be, is mentioned in Sect 6. below.

5.3 Non-standard Gravitations and Cosmologies

The possibility of non-gravitational forces had been mentioned briefly at the 1961 conference (cf. de Vaucouleurs 1960 and Hogan and White 1986 for a more recent version.) Non-Newtonian (or non-relativistic) gravity can be equivalent, and appears quantitatively first in the work of Finzi (1963),who marshaled evidence for M/L values increasing outward in the Milky Way, and so suggested that the deviation was in the direction of G becoming stronger than $1/R^2$ at large distances. Currently-viable variants of alternative gravity belong to a later period, while non-zero Λ never quite disappears from the astronomical recipe book (McCrea 1971).

In case any reader is shouting to himself "magnetic fields," we remind you that these constitute a highly relativistic fluid. Thus if they are to confine anything, they must be firmly anchored to gravitating matter, which must be sufficient to contain the fields themselves as well as whatever they originally set out to hold on to.

5.4 Gravitational Lensing (and Deflection of Light)

The early literature of this topic is much richer than is generally recognized. Trimble (2001) cites 8 papers in German and 10 in English from before 1961 and 8 in Russian and at least 15 in English from 1961-74. Among the items more or less understood were:

- There is no precise analog with bending (etc.) of light by material lenses, because it is not possible to have a region of constant index of gravitational refraction, but rays passing just on either side of the sun would meet outside the orbit of Neptune, making it an F = 59,600 system.
- The Newtonian/particle theory of light calculation was done by Soldner, Cavendish, and others before 1900, besides Michell and Laplace, and, of course, Einstein in 1907-08.
- The extra factor of two bending in General Relativity constitutes a test of the theory.
- The possibility of both rings and multiple images in what we now call strong lensing was pointed out in the 1920s (by Oliver Lodge and O. Chowlson, of St. Petersburg but writing in German), and Edwin Frost at Mt. Wilson apparently looked for star-star lensing in 1923. (He didn't find any.)
- Rudi W. Mandl drove both the interwar calculations by Einstein and by Zwicky.
- Zwicky (1937) understood that gravitational lensing could measure galaxy masses, magnify the images of distant sources, test general relativity, and was more likely for galaxies than for stars.
- H.N. Russell imagined what an inhabitant of a planet orbiting Sirius B might see, including lensed arcs from extended sources.
- Other well-known people dabbled in the territory (Charles Darwin, the grandson of the one you know, Ya. B. Zeldovich, James E. Gunn, P.J.E. Peebles, Philip Morrison and his students…)
- Weak lensing was also calculated with increasing precision, though generally viewed as a noise source – like "a shower door", imposing fundamental limitations on the accuracy with which you can measure angular diameters of distant galaxies – and it was shown that the optical depth of the universe to weak

lensing by any population is roughly the same as the fraction of closure density that population contributes.

- The effects of galaxies and clusters on the cosmic microwave background include lensing effects
- Optical depths for star-star lensing were calculated by Lieves (1964) who found magnifications of 2000 possible, but an optical depth in the Milky for even a factor 10 only 10^{-9}.

The first paper likely to be mentioned in recent discussions of gravitational lensing is, however, that of Refsdal (1964). This was a true “sleeping beauty” paper in the language of the community that analyzes citation rates. Almost ignored at the time of publication and for at least 15 years afterwards, it is now widely known, perhaps most because the author pointed out you could use the multiple lensed images of variable QSOs or quasars (with known redshifts for both source and lens) to measure the Hubble constant. In the process, you also get a mass for the lens, as Zwicky (1937) promised.

Strong and weak lensing produce multiple images and distortions respectively. The phrase microlensing can describe either variability in a strongly-lensed source due to individual stars or other lumps in the lensing galaxy or star-star lensing in and around the Milky Way. These things have now all been seen, starting only in 1979 (Walsh et al. 1979), though the Barnothys (1968) had suggested that all QSOs and quasars were lensed and intrinsically only as bright as Seyfert galaxies. The phenomena come into their own relative to dark matter in three ways: tests for the nature of the galactic halo DM from microlensing searches; detection of dark halo substructure in galaxies that lens, and measurements of cluster masses to compare with the numbers from X-ray data and the Virial theorem.

5.5 The Cosmic Microwave Background and its Remarkable Isotropy

The discoverers of the CMB do not need any more citations (but for a whole book about the subject, see Peebles et al. 2009). This radiation is important for dark matter because it is so smooth across the sky that only non-baryonic DM can build the galaxies we see in the time allowed.

Even the first papers noted that the temperature or intensity did not seem to vary on any angular scale, with an upper limit of 10%. That limit shrank monotonically with time, the dipole due to our motion popping out at a part in 10^3 in about 1970 (Conklin 1969, Henry 1971). Meanwhile, as it were, astronomers (e.g. Rees 1971) were trying to form galaxies and clusters from the gradual growth of density fluctuations established in the very early universe. Now the problem is as follows:

From recombination to the present is z = 1000 to 0, meaning that linear perturbations can grow only about a factor of 10^3. Thus $\Delta\rho/\rho = 1$ now requires $\Delta\rho/\rho\ 10^{-3}$ then. But, for adiabatic fluctuations $\Delta T/T = (1/3)\ \Delta\rho/\rho$, since photon number density scales as $T^{1/3}$. Thus, for galaxies and clusters formed entirely from baryons, we predict that the radiation, streaming freely after recombination should be lumpy on the sky at a level of 3×10^{-4}. It is not. Zeldovich (1972) considered the possibility of isothermal fluctuations, but the underlying physics to produce these was problematic.

Dark matter made of baryons does not help, but DM that interacts with photons only gravitationally could be a galaxy-saver, predicting CMB fluctuations of parts in 10^5 rather than parts in 10^4. Density perturbations can grow before recombination to larger amplitudes and gas can stream into the previously-assembled DM halos after recombination to become galaxies and all. The most prominent DM candidates belong to a later time frame, and the tip of the relevant CMB fluctuations were seen only with the launch of the COBE satellite (Smoot et al. 1992), with the ones directly connected to current structures belonging to the very recent epochs of WMAP and some preceding balloon-borne telescopes (Vol. 6)

5.6 The First Non-Baryonic Candidates

At least five of these appear in the literature before 1974. All are now thought to be likely, but not dominant, contributors. The existence of neutrinos was generally accepted from the mid 1950s and their masses understood to be zero or small. The laboratory limits were, however, not very tight, and Gershtein & Zeldovich (1966) realized that one could do better simply by insisting that the universe not be strongly over-closed by them. Their limit was about 400 eV for a single type of neutrino or anti-neutrino, and Ω less than 10. A more stringent limit on Ω and equal weight given to electron and mu neutrinos and their antiparticles led Cowsik and McClelland (1972) down to less than 20 eV, but within the range that could close the universe. Laboratory measurements of the mass differences among the three flavors now known are a good deal smaller and make them all probably less than 1% of Ω.

A second possibility, from the same Russian school in the same year (Zeldovich & Novikov 1966) was primordial black holes. This means ones formed before the epoch of nucleosynthesis, so that baryons could have gone into them without disturbing the production of helium and deuterium (sect. 5.9). Again the authors were not primarily thinking of PBHs as dark matter, and indeed they are not if there is gas around to be accreted and to radiate as it goes. In addition, BHs can bend light coming to us from sources behind them (gravitational microlensing for BHs of planetary and stellar masses), and for asteroidal masses around 10^{15} g perhaps boil away (Hawking radiation), while large ones (e.g. 10^6 $M_\odot$) in galaxies will wreak havoc with star clusters, giant gas clouds, and even whole disks. Large numbers of 10^8 $M_\odot$ or larger black holes would lens QSOs in a way not seen. These five processes enable the exclusion of most, but perhaps not quite all, masses of black holes as dark matter candidates.

In a short, prescient paragraph, Sciama (1971, p. 129) mentions what he calls "missing material," saying that the individual faint stars, rocks, neutrinos, or gravitational waves would not yet have been observed. Like PBHs, gravitational waves of most wavelengths would have revealed themselves in the interim (for instance by wiggling the clocks we call pulsars), but there is possibly still some residual phase space. The longest wavelength gravitational waves are likely to be those left from the early universe ("primordial"), discussed by Grischuk, (1974).

Yet another idea from fundamental physics is that of various space-time singularities, now called monopoles, strings, domain walls, and textures. These can be produced by spontaneous symmetry breaking during phase transitions (Krizhnits & Linde 1972), and it was immediately clear that one does not want very many of any of them around (Zeldovich et al. 1974).

5.7 Rotation Curves Again

For whatever reason, this is the length scale on which dark matter first acquired mainstream acceptance, and even non-technical discussions of the topic typically begin by saying that the outer regions of galaxies would fly apart without the extra gravitational force from dark matter. The data, however, belong only just barely to the 1961-74 period, and then only for M31, with improved optical data (Rubin & Ford 1970) and 21 cm measurements extending further out (Roberts & Rots 1973; Roberts & Whitehurst 1975). The key point is that the rotation speeds do not decline outside the optically bright portions of galaxying, implying that M/L is rising outward. Freeman (1970) was nearly alone in appreciating that the phenomenon might be widespread and would be important.

Again somewhat inexplicably, the optical rotation curves have been more widely recognized, so that, for instance, the conference proceedings edited by Kormendy & Knapp (1985) mentions Rubin on 37 pages and Roberts on only 4. Trimble (1987 Sect. 2.2) is similarly guilty, though perhaps only at the 2:1 level only.

5.8 Pure Theory: Disk Instabilities

A thin, self-gravitating disk is unstable to bar formation and various other distortions (sometimes seen in accretion disks as well as galaxies). Ostriker & Peebles (1973) are generally credited for pointing out the instability and for showing that a spheroidal halo of comparable mass could stabilize them. Infrared imaging has shown that there are more barred spirals than were thought at that time (Casasola et al. 2008).

5.9 Big Bang Nucleosythesis

This topic is relevant because, by 1974, it could just about rule out a universe closed by baryons. George Gamow and his younger colleagues had tried to make the entire range of chemical elements from neutron matter in an early universe (heated by the decaying neutrons). This fails most conspicuously because there are no stable nuclides with A = 5 or 8. Thus the next stage was "the early universe made hydrogen and helium, but Burbidge, Burbidge, Fowler, and Hoyle (1957) made all the rest." All four were, to varying degrees, steady state supporters, thus it is at least curious that the latter two participated in one of the early quantitative calculations of how much H and He you should get from a Big Bang (Wagoner, Fowler, & Hoyle 1967). Wagoner (1973) improved the calculations, allowing for dependence on expansion time scale, neutron half life and other uncertain nuclear physics, and baryon density, among other variables.

He/H is rather insensitive to most of those variables, so that the discovery of interstellar deuterium at a level D/H = 1.5×10^{-5} (Rogerson & York 1974) helped a good deal to tie things down. Thus Gott, Gunn, Schramm, and Tinsley (1974) could say that a cosmic (baryon) density in excess of about $\Omega = 0.15$ would yield more helium and less deuterium than we see around us. What you get depends separately on the baryon density (a matter of how easy it is for particles to find each other) and the total density (the expansion time available for them to do it) as pointed out, perhaps first, by Shvartsman (1969).

Gott et al. (1974) also invoked the M/L values of large clusters (perhaps 200 in solar units) and the ages of population II stars and radioactive elements vs. the reciprocal of the then-favored Hubble constant to conclude that the universe is open. These three items were all correct in the sense of revealing that the total density of matter in any form is considerably less than $\Omega = 1$. Though the paper was published on 1 April, Λ was not mentioned, but Gunn & Tinsley (1975) had a go at an $\Omega = 1$ universe with large values of both ρ and Λ which did not catch on at the time.

In the ensuing 35 years, holding by the requirement that the early universe must make the proper amounts of H, H^2, He^3, He^4, and Li^7 has helped to rule out a wide range of non-standard cosmologies, including strong anisotropies and decaying particles, that belong to Vol. 6 of this series.

5.10 The 1974 Papers

Jaan Einasto (2001) has described in some detail the logical process that led up to the first of the two watershed papers. His own work on the largest available sample of binary galaxies was key, as was the insistence by Zeldovich that he and his colleagues submit to a major journal (Einasto, Kaasik, & Saar 1971). Their key points were "The mass of galactic coronae halos exceeds the mass of populations of known stars by one order of magnitude, as do the effective dimensions" and "The mass-luminosity ratio rises to f ≈ 120 for elliptical galaxies. With H = 50 km/sec/Mpc, this ratio for the Coma cluster is 170." A careful reading of their paper suggests that they had meant to include a table or graph of M/L vs. length scale, which somehow got left out (Nature papers were still very short in those days).

Einasto et al. were just in time for the second paper (Ostriker, Peebles, and Yahil 1974) to cite their preprint. The latter concluded "Currently available observations strongly indicate that the mass of spiral galaxies increases almost linearly with radius to nearly 1 Mpc…and that the ratio of this mass to the light within the Holmberg radius, f, is ~ 200 ($M_\odot/L_\odot$). They show a graph of M/L v.s length scale which would extrapolate to the equivalent of closure density (M/L ~ 1000, remember) at R = 3000 Mpc, the Hubble radius. Over the years, the Ostriker paper has been cited about twice as often as the Einasto paper. It would be interesting to chase down a large sample of the citations and decide whether the main driver is simply American chauvinism.

Very soon after, Ozernoy (1974) pointed out that, within rich clusters, the dark matter must mostly belong communally to the cluster and not to the individual galaxies, first because the galaxy separations (at least in cluster cores) were smaller than the necessary DM halo sizes, and second because there is not as much luminosity segregation as would be expected from mass segregation if those hefty monsters were interacting gravitationally.

A first counterblast to this new dark matter paradigm came from Burbidge (1975) emphasizing, as Holmberg had done in 1961, selection effects, observational errors, and other uncertainties. He remained unreconciled to what is now conventional dark matter at least until 2001 (Narlikar, 2001, Sect. 13).

6. Dark Matter and Standard Models Since 1974

Practicing scientists will normally put the cut between history and current events at the time when they started reading the literature for themselves, probably early in graduate school. This was about 1965 for the present writer, but could easily be 40 years later for some readers (and perhaps as much as 20 years earlier for very few others). Some time in that 40 years, pervasive, at-most-weakly interacting, dark matter became part of customary astronomy and cosmology. This section is intended to bridge the gap from the previous ones to the considerations of modern cosmology that will appear in Volume 6. The main topics are the impact of inflation, the cosmological constant yet again, dark candidates, structure formation, residual doubts, and dark matter alternatives. These are not arranged chronologically, but perhaps in the order you might meet them in an introductory textbook.

6.1 The Impact of Inflation

The general idea is a very early (before 10^{-23} seconds or so) epoch of exponential cosmic expansion, followed by reheating to the high temperatures required for nucleosynthesis, thermalization of neutrinos, and so forth. Reserving the name Big Bang for that latter period would save a great deal of fuss and bother, because we have considerable observational evidence that it actually happened. Inflation solves several problems called causality, monopoles, flatness, large scale homogeneity, and origin of small scale inhomogeneities (Chapter X of Vol. 6) that had not bothered most astronomers until the particle physicists told us about them (compare telephone solicitations for aluminum sliding and comprehensive telephone plans).

From the point of view of dark matter, the key property of inflation is that it guarantees that the universe will be very nearly flat (k=0) and have the closure density in the sum of all forms of matter plus cosmological constant, because a very large increase in the scale parameter R(t) in the relevant equation drives the k term to zero:

$$H^2 = \frac{8\Pi G\rho_{in}}{3} - \frac{k}{R^2}$$

where H is the Hubble constant, G is the constant of gravity, ρ_{in} is the energy density of the scalar field invoked to drive the inflationary (exponential) expansion, and c = 1 in the units loved by cosmologists who don't build apparatus. Various forms of matter and radiation dominate Λ until rather recently, if Λ is truly a constant. The key papers normally cited are Guth (1981) and Linde (1982). A good didactic presentation is that of Kolb & Turner (1988, Chapter 8).

Thus the idea of inflation drove astronomers back to thinking about a critical-density universe, initially closed by some form of (dark) matter and later by some form of Λ, cosmological constant, dark energy, or quintessence.

6.2 The Final Resuscitation of Λ

The cosmological constant has hung around the fringes of the universe since 1917. Einstein (1917) dropped it from his later cosmological papers, but the evidence that he claimed it as his greatest blunder is third-hand. Zeldovich (1968) pointed out that Λ could be thought of as a vacuum field energy density, for which the Lamb shift, Casimir effect, etc. provide independent evidence of a sort. But its re-incorporation into mainstream cosmology was driven by observations. Kofman et al. (1993) presented perhaps the first "Λ CDM" universe, based on COBE measurements of the CMB anisotropy, large scale clustering, and ages of things. Their Λ was about 0.8 of the critical density, with the rest in some combination of cold dark matter and baryons. By 1995, Peebles (1995) is ready to say that "the last column [of Table I] assumes at the present epoch the mass density is subdominant to only one significant term, Λ or space curvature," the latter meaning an open universe. And at an IAU Symposium (S183, Cosmological Parameters and Evolution of the Universe), held as part of the General Assembly in Kyoto in 1997, a panel discussion ended with a sort of vote in which the majority of the participants endorsed something like the present consensus model, with H around 70 km/sec/Mpc, 5% of the closure density in baryons, another 20% or so in dark matter, and the rest in cosmological constant. Unfortunately, the discussion does not appear in the symposium proceedings, and three of the panel members who made up part of the majority were co-opted after the program book was printed (J.P. Ostriker, the present author, and one other), so that the record is only in the memories of the participants and any surviving notes they might have taken at the time.

The yearly review, Astrophysics in 1997 (Trimble & McFadden 1998) was the first in a series of 16 to record a majority of papers favoring non-zero Λ. Then, starting in 1998 came the supernova and WMAP data that completed the process of Λ incorporation (Spergel et al. 2007).

6.3 Candidates: Dark, Darker, and Darkest

Sect. 5.6 mentioned neutrinos, black holes, gravitational radiation, topological singularities (useful really only as seeds for galaxy formation); and the main baryonic candidates (very faint stars, stellar remnants, and gas in elusive density/temperature regimes) are left from Sect. 1. Many more recent ones have come from considerations in elementary particle physics. First of the two most generally invoked are weakly interacting super partners of known bosons, given names ending in "ino" (like photino and gravitino) or, more generally, WIMPs (Weakly Interacting Massive Particles, the name introduced deliberately by Steigman and Turner, 1984). The basic scheme was laid out by Lee and Weinberg (1977), showing that likely particle masses (GeV or more) and cross sections (10^{-42} cm^2 or less) indeed yield an appreciable, but nearly undetectable, cosmic density. The other entity, axions (Peccei & Quinn 1977), is co-eval, at least in the literature, and is also a possible form of cold dark matter, non-relativistic at the epoch of galaxy formation, because they form out of a Bose-Einstein condensate, despite their masses being much less than an eV.

And then there are all the rest. The total number tabulated by Trimble (2005) is about 75, and the list is by no means exhaustive. Trimble et al. (2007) present 23, not all new. Some arise from front-line physics and could be part of a unified dark sector. Some favorites are SuperWIMPS (meaning even smaller cross sections), WIMPzillas with masses of 10^{15} GeV, and Kaluza-Klein particles. Some had brief periods of glory that quickly faded, like self-interacting dark matter (larger cross sections but only for particles encountering each other) and decaying dark matter with mass = 27 eV. And some are just odd, for instance DAEMONS (DArk Electric Matter Objects). These are particles for which thermodynamic time runs backwards, preventing electromagnetic interactions.

The opposite of a WIMP is, of course, a MACHO, a MAssive Compact Halo Object in the Milky Way halo that could gravitationally lens stars in the Large Magellanic Cloud or in the disk that could lens stars in the bulge. The lensing is seen, and dividing up the lenses among known faint stars in the Milky Way and LMC and expected MACHOs (white dwarfs, neutron stars, black holes, other very faint entities) still presents some problems (Torres 2008). Whatever they are, however, they are not most of the mass in our halo. The original MACHO project completed its intended surveys (Alcock et al. 2003), though some related ones (called OGLE, ASAS, etc.) continue and are now finding large numbers of planets transiting their (non-lens) host stars and a few lenses with planets as well.

6.4 Galaxy and Cluster Formation with Dark Matter

No sooner had Lee and Weinberg (1977) codified the idea of supersymmetric partner particles than astrophysicists began incorporating them, as cold dark matter, into their models of structure formation (Gunn 1977, White and Rees 1977). This was, from the beginning, biased cold dark matter, because the results of calculations resembled the real masses and spacings of galaxies only if the galaxies were restricted to the highest peaks of DM density, so that $\Delta L/L = b \Delta M/M$, with the bias factor b = 1.5 – 2.5. Blumenthal et al. (1984) is generally regarded as the definitive demonstration that biased CDM was beginning to solve "the galaxy formation problem," at least on scales up to about 10 Mpc. And see Solovevo and Starobinskii (1985) for the view from further east.

On larger scales, however, the simulations did not produce nearly as many of the 30-100 Mpc voids, clusters and deviations from uniform Hubble flow as are seen in the real world. Let Bertschinger (1991) stand for very many papers reporting this problem, and Hamilton et al. (1991) for a comparable number pointing out that almost anything added – including baryons, hot dark matter, topological defects or other seeds for galaxy formation, a complex spectrum of initial perturbations, or, as we now recognize, non-zero Λ – would improve the situation.

Where was the hot dark matter while all this was going on? Lurking on the sidelines from the time when Tremaine & Gunn (1979) pointed out that the exclusion principle did not allow enough quantum states for neutrinos with realistic masses to bind the smallest dwarf galaxies with their shallow potentials. A few years later, HDM was to be found as an adjunct for producing the largest scale structures (Umemura & Ikeuchi 1986, Doroshkevich 1984). Recent work on galaxy formation is happy with Ω (HDM) = 0.006, about the number expected from the known neutrino mass differences (Allen et al. 2003).

Sterile neutrinos are ones that do not participate in either big bang nucleosynthesis or neutrino oscillations under current conditions. If they close the universe, they could be massive enough to be Warm Dark Matter (semi-relativistic) at the epoch of galaxy formation. Knebe et al. (2008) have reconsidered the possible role of WDM for galaxies.

Virtually all 21st century simulations of the assembly of galaxies and clusters, however, make use of a standard ΛCDM mix of ingredients. The Millenium Simulation (Springel et al. 2005) was the one to beat, and to use, a few years ago, and has indeed probably now been beaten by somebody, after a few additional "Moore's Law quanta". There are two issues: first the need to incorporate as many physical processes (including baryon ones) as possible, and, second, the drive for ever-larger numbers of particles in the calculation, so as to be able to resolve structures of $10^6 M_\odot$ (dwarf galaxies; young globular clusters) or less. Incidentally, confidence in the main outlines has grown to the point where the presence or absence of dark matter is used to distinguish very small galaxies (with) from large star clusters (without, Mieske et al. 2008, Fellhaver 2008). The distinction between a large bathing machine and a very small second class carriage is left as an exercise for the reader.

6.5 Residual Doubts

Well into the "consensus era," a small number of brave observers continued to report data sets that were, they said, more nearly, or at least as, consistent with $\Omega_M = 1$ or Ω (total) small (Trimble & Aschwanden 2003, Sect. 12.2). On the other hand, an earlier claim that dark matter should not be taken seriously, because rich clusters yielded three different numbers from the velocity dispersion, X-ray data, and weak lensing was resolved. Cohen & Kneib (2002), for instance looked at RX J1347-1145 in all three ways and concluded that the Virial mass was smallest, the X-ray mass largest, and lensing came in the middle. They suggested that the cause might be a merger in progress, which has shocked the gas (making it hotter than it ought to be) while the two separate velocity dispersions have not yet had a chance to discover that they belong to a single, more massive cluster. Generally, non-thermal pressure support (shocks, magnetic fields, turbulence) will yield a too-large X-ray mass; substructure and non-members can either enlarge or contract the velocity dispersion; and lensing is not the desired gold standard, because the distortions are subtle enough you have to postulate some spherical or other simple model for the lensing cluster to carry out the analysis.

At the other extreme from these almost-mainstream reservations come the sorts of items you may well have received as part of a mass postal or e-mailing. The most recent on my desk comes from Tsiganov & Tsiganov (2009) and expresses the view that conventional gravitation physics went wrong somewhere around the time of Galileo and Huygens. The illusion called dark matter is, they say, one of very many consequences of this mistaken view of gravity. The authors are at the University of Economics and Law in Ukraine, but other papers almost as remarkable have come from departments of physics, astronomy, and engineering.

6.6 Dark Matter Alternatives

Now that Λ is part of the standard universe, this has come to mean theories of gravity other than general relativity and cosmologies other than Friedman-Robertson-Walker-Lemaitre which wholly or partly replace dark matter. Many of these have short half lives, are supported only by their originators, and have not been brought face-to-face with the full range of relevant data. This range includes solar system tests, the changing orbits of binary pulsars, the redshifts of X-ray lines from X-ray binaries with black hole accretors, and rotation of gas and velocity dispersions in galaxies, as well as structure formation, weak, strong, and microlensing, and the evolution of merging galaxy parts. Trimble (2005) summarizes half a dozen such theories. Vekeschagin & Yegorian (2008) review a larger, partly-overlapping set.

Closest to being fully-worked-out alternatives are MOND (MOdified Newtonian Dynamics) and its relativistic extension (Milgrom 2007, Bekenstein 2006, and earlier papers referenced therein). The general idea is that of a minimum possible gravitational acceleration, below which gravity turns over to a 1/R rather than $1/R^2$ Law. Of at least equal importance from the point of view of history of science is that a sizable number of astronomers, astrophysicists, etc. (at least a dozen) have taken the idea seriously enough to explore its consequences (S.M. McGaugh, pers. comm. 2009 kindly provided a list of the men (and a very few women) of MOND.) Thus, while a typical month of a major journal is like to have 5-10 or more papers exploring the properties and consequences of dark matter or assuming its existence for some other purposes like data analysis, there will also usually be one or two MOND papers. I happened to look at the June 2008 issues of MNRAS and the June 2009 issues of ApJ. Many of these current authors are younger than the originators of the idea, suggesting it will have a long half-life. In some cases, dark matter (conceivably sterile neutrinos) is needed even with MOND (Richtler et al. 2008).

6.7 From the Standard Hot Big Bang to Consensus Cosmology

Both the Big Bang and inflation have passed a number of tests that they might have failed, coming from data on primordial abundances, CMB brightnesses and temperature isotropy on large scales plus small fluctuations on small scales, measurements of ages of stars and radioactive elements vs. the Hubble constant, distant supernovae, and so forth. The present belongs to WMAP and its implications (Komatsu et al. 2009 on the 5th year data release).

Some unsolved problems, largely connected with small scale structures, remain. Simulations tend to produce cuspy centers for galaxies and clusters, while we see more nearly isothermal cores (Siman & Geha 2007), and also more substructure in the halos of big galaxies than are seen as satellites (Abdelqudev & Melia 2008; Madau et al. 2008). Most researchers coming across these problems seem to think that better calculations, with more baryon physics and better mass resolution, are likely to solve them.

In summary, we can probably say that the question of the existence of dark matter has been subsumed by the consensus cosmology model. Its nature is a different story, given that there are very many candidates that cannot entirely be excluded and that the right one is perhaps still not in the inventory. How can progress be made? More and better calculations and observations of large scale structure and CMB distortions may be part of the story, but the happiest event would be some form of detection. There are three possibilities: observation of decay or annihilation products (photons, neutrinos, leptons) coming from the Galactic halo; capture in laboratory detectors; or production by particle accelerators. Several claims of the first two have appeared in recent years with at least partial refutations, addressed in Hooper & Boltz (2008) on DM and Caldwell and Kamionkowski (2009) on DE.

Acknowledgements

I am most grateful to Gerard Gilmore for the invitation to write this review and to Alison Lara for keyboarding it. The astronomers, some no longer with us, who have contributed to and influenced my thinking about dark matter range from Abell (Geroge Ogden) to Zwicky (Fritz), with special extra thank yous to Vera Cooper Rubin and Daniel Magnus Popper for insights on “how things used to be.”

REFERENCES

Abdelgudeu, M. & Melia, F. 2008, MNRAS 388, 1869

Adams, J.C. 1847. Mem. RAS 16

Alcock, C. et al. 2003. ApJ 598, 597

Allen, S.W. et al. 2003. MNRAS 346, 593

Babcock, H.W. 1939, Lick Observatory Bulletin 19, 41

Barnothy, J. & Barnothy, M.F. 1968. ApJ 174, 477

Bekenstein, J.D. 2006. Contemporary Physics 47, 387

Bertschinger, E. 1991. In S.S. Holt et al. eds. APS Conf. Proc. 272, p. 297, After the First Three Minutes

Bessel, F.W. 1845a. AN 22, p. 145, 169 & 185

Bessel, F.W. 1845b, MN 6, 136

Blumenthal, G., Faber, S., Primack, J. & Rees, M. 1984, Nature 311, 587

Bondi, H. & Gold, T. 1948. MN 108, 252

Burbidge, E.M. & Burbidge, G.R. 1961. AJ 66, 541

Burbidge, G.R. 1975. Astrophys. J. 198, L7

Burbidge, E.M., Burbidge, G.R., Fowler, W.A., & Hoyle, F. 1957. Reviews of Modern Physics 21, 547

Caldwell, R. & Kamionkowski, M. 2009. Ann. Rev. Nucl. & Part. Sci.

Casasola, V. et al. 2008. A&A 490, 61

Cicaire, I. et al. 2008, AJ 135, 2038

Clarke, T.E. et al. 2003. ApJ 601, 798

Cohen, J.G. & Kneib, J.-P. 2002, ApJ 572, 524

Conklin, E.K. 1969, Nature 222, 971

Cowsik, R. & McClelland, J. 1972. PRL 29, 669

Da Rocah, C. & Mendes de Oliveira, G. 2005. MN 364, 1069

de Vaucouleurs, G. 1958. ApJ 128, 465

de Vaucouleurs, G. 1960. ApJ 131, 585

Doroshkevich, A.G. 1984. Soviet Astronomy AJ 28, 253

Einesto, J. 2001. In V. Martinez et al. eds. Historical Development of Modern Cosmology, ASP Conference Series 252, 85

Einesto, J., Kaasik, A. & Saar, E. 1974. Nature 250, 309

Einstein, E. 1917. Sitz. König. Preuss. Akad. Wissenschaften, Berlin, 142-152

Fasenko, R. 1985. Astrofizika 20, 495

Fellhaver, M. 2008. MN 385, 1095

Felten, J.R. & Morrison, P. 1966. ApJ 146, 686

Field, G. & Henry, R. 1964, ApJ 140, 1002

Finzi, A. 1963. MN 127, 21

Freeman, K.C. 1970. Astrophys. J. 160, 811

Gershtein, S.S. & Zeldovich, Y.B. 1966. Journal of Experimental and Theoretical Physics Letters, 4, 175

Ghez, A. et al. 2008. ApJ 689, 1044

Giacconi, R. et al. 1962, PRL 9, 439

Gott, J.R., Gunn, J.E., Schramm, D.N. & Tinsley, B.M. 1974, ApJ 194, 543

Gould, R. & Burbidge, G.R. 1963, ApJ 138, 969

Grischuk, L.P. 1974. JETP 40, 409

Gunn, J.E. 1977. ApJ 218, 592

Gunn, J.E. & Peterson, B.A. 1965. ApJ 142, 1637

Gunn, J.E. & Tinsley, B.M. 1975. Nature 257, 454

Gursky, E. & Schwartz, D.A 1977, ARA&A 15, 541

Guth, A.R. 1981. PRD 23, 347

Halley, E. 1718, Philosophical Transactions, Royal Society for 1718

Hamilton, A.M.S. et al. 1991. ApJ 374, L1

Henry, P.S. 1971. Nature 231, 516

Hockey, T. et al. Eds. 2007. The Biographical Encyclopedia of Astronomers, Springer

Hogan, C. & White, S.D.M. 1986, Nature 321, 575

Holmberg, E. 1937. Lund Observatory Annals 6, 173

Holmberg, E. 1954. Lund Medd. Series I, 186

Hooper, D. & Boltz, F.A. 2008. Ann. Rev. Nucl. & Part. Phys. 55, 293

Hoskin, M. Ed. 1997, Cambridge Illustrated History of Astronomy, Cambridge Univ. Press

Hoyle, F. 1948. MNRAS 108, 372

Hoyle, F. 1959. In IAU Symp. 9, Ed. R.N. Bracewel, Stanford Univ. Press, Paris Symposium on Radio Astronomy

Hoyle, F. 1963. ApJ 137, 993

Hubble, E.P. 1934, ApJ 76, 44

Jeans, J.H. 1922, MNRAS 82, 122

Kahn, P. & Woltjer, L. 1959. ApJ 130, 105

Kapteyn, J.C. 1922. ApJ 55, 302

Knebe, A. et al. 2008. MN 386, 1029

Kofman, L.A. Gnedin, N.Y. & Bahcall, N.A. 1993. ApJ 413, 1

Kolb, E.W. & Turner, M.S. 1990. The Early Universe, Addison Wesley

Komatsu, E. et al. 2009. ApJ Suppl. 180, 330

Kormendy, J. & Knapp, J. Eds. 1985. IAU Symp. 117, Dark Matter in the Universe, Dordrecht: Reidel

Kragh, H. 1996. Cosmology and Controversy. Princeton University Press

Kragh, H.S. 2007. Conceptions of the Cosmos. Oxford Univ. Press

Kriznits, D.A. & Linde, A. 1972. Physics Letters 42B, 471

Kuhn, T. 1962, The Structure of Scientific Revolution, Univ. Chicago Press

Kunth, R. 1952. Zeitschrift fuer Astrophysik 28, 234

Kuzmin, G.G. 1952. Tartu Astronomical Observatory Publications 32, 211

Lalleman, A. 1960. PASP 72, 76

Laplace, P.S. de 1796. Exposition du system du Monde, Vol II, p. 305, Paris

Lee, B. & Weinberg, S. 1977. PRL 38, 2237, PRL 39, 165

Leverrier, U.J.J. 1847. AN 25, 53

Leverrier, U.J.J. 1848. AN no. 580

Lieves, S. 1964. PR 133, B858

Lindblad, B. 1926. Uppsala Medd. No. 3

Linde, A.D. 1982. Physics Letters 108B, 389

Lohman, W. 1956. Zeitschrift fuer Physik 144, 66

Madau, P. et al. 2008. ApJ 679, 1261

Mayall, N.U. & Aller, L.H. 1960. PASP 52, 278

McCrea, W.H. 1971. Quarterly Journal of the Royal Astronomical Society 12, 40

McVittie, G.C. Ed. 1961. Problems in Extragalactic Research, IAU Symp. 15, New York: Macmillan

Michell, J. 1767. Philosophical Transactions Royal Society, p. 243

Michell, J. 1784. Phil. Trans. RS 74, 35

Mieske, S. et al. 2008. A&A 487, 921

Milgrom, M. 2007. ApJ 667, L45

Narlikar, J. 2001. in V. Martinez et al. Eds. Historical Development of Modern Cosmology, ASP Conf. Ser. 252, p. 175

Neyman, J. Page, T. & Scott, E. 1961 Eds. AJ 66, 537 (Conf. on Instability of Clusters of Galaxies)

Newcomb, S. 1906. Sidelights on Astronomy: Essays and Addresses, NY: Harper & Brothers

Oort, J.H. 1927. BAN 3, 275

Oort, J.H. 1932. BAN 6, 249

Oort, J.H. 1940. ApJ 91, 227

Oort, J.H. 1965. in A. Blaauw & M. Schmidt Eds. Galactic Structure (Vol V of the Kuiper compendium) Univ. of Chicago Press

Opik, E. 1915. Bull. de la. Soc. Astr. de Russie 21, 150

Opik, E. 1922. ApJ 55, 406

Ostriker, J.P. Peebles, P.J.E. 1973. ApJ 186, 467

Ostriker, J.P.& Peebles P.J.E. & Yahil, A. 1974. ApJ 193, L1

Ozernoy, L.M. 1974. A. Zh. 51, 1108 (Soviet Astronomy AJ 18, 654)

Page, T.L. 1952. ApJ 116, 23

Page, T.L 1956, ApJ 136, 685

Page, T.L. 1960. ApJ 132, 910

Peccei, R.D. & Quinn, H.R. 1977. PRL 38, 1440

Peebles, P.J.E. 1995, in V. Trimble & A. Reisnegger Eds. Clusters, Lensing, and the Future of the Universe, ASP Conf Ser. 88, p. 1

Peebles, P.J.E. et al. 2009. Eds. Finding the Big Bang, Princeton Univ. Press

Pigott, E. 1805. Phil. Trans. Roy. Soc. 95, 131

Pooley, D. et al. 2008. ApJ 697, 1892

Read, J.I. et al. 2006, MNRAS 371, 885

Read, J.I, et al. 2008. MNRAS 389, 104

Rees, M.J. 1971 in B.K. Sachs, Ed. Italian Physical Society, Proceedings of International School of Physics, “Enrico Fermi”, Course 47, General Relativity and Cosmology, NY: Academic Press

Refsdal, S. 1964. MNRAS 128, 295, & 307

Richter, P. et al. 2006. A&A 445, 827

Richtler, T. et al. 2008. A&A 478, L23

Roberts, M.S. & Rots, A.H. 1973. A&A 26, 483

Roberts, M.S. & Whitehurst, R.N. 1975. ApJ 201, 377

Rogerson, J.G. & York, D.G 1974. ApJ 186, L95

Rubin, V.C. & Ford, W.K. 1970. ApJ 159, 379

Russell, H.N., Dugan, R.S., & Stewart, J.Q. 1926. Astronomy, Boston: Ginn & Co.

Schaeberle, J.M. 1896, AJ 17, 37

Schmidt, M. 1956. BAN 13, 14

Schwarzschild, M. 1954. AJ 59, 273

Sciama, D. 1971. Modern Cosmology. Cambridge Univ. Press

Shvartsman, V.F.1969. JETP 9, 184

Siman, J.D. & Geha, M. 2007. ApJ 670, 313

Slipher, V.M. 1914. Lowell Observatory Bulletin, 2, 65

Smith, S. 1936. ApJ 83, 23

Smoot, G. et al. 1992. ApJ 396, L1

Solovevo, L.V. & Starobinskii, A.A. 1985. Sov. Astron. AJ 29, 367

Spergel, D.N. et al. 2007. ApJ Suppl. 170, 377

Springel, V. et al. 2005. Nature 435, 629

Steigman, G. & Turner, M.S. 1984. Nuclear Physics B252, 73

Tinsley, B.M. & Larson, R.B. 1977. Eds. The Evolution of Galaxies and Stellar Populations. Yale University Observatory

Tolman, R.C. 1934, Relativity, Thermodynamics, and Cosmology, Oxford Univ. Press

Torres, S. 2008 A&A 486, 427

Tremaine, S. 1987 in Kormendy & Knapp Eds. 1987, p. 547

Tremaine, S.D. & Gunn, J.E. 1979. PRL 42, 407

Trimble, V. 1987. Annual Reviews of Astronomy & Astrophysics 25, 425

Trimble, V. 1988. in R. Bertotti et al. Eds. Modern Cosmology in Retrospect, Cambridge Univ. Press. P 355

Trimble, V. 1995. in S.S. Holt & C.L. Bennett Eds. Amer. Inst. Phys. Conf. Proc. 336, 57

Trimble, V. 1996. PASP 108, 1073

Trimble, V. 2001. in T.G. Brainerd & C.S. Kochanek Eds., Gravitational Lenses, Astron Soc. Pacific Conf. Ser. 237, p. 1

Trimble, V. 2005 in M.M. Shapiro et al. Eds. Neutrinos and Explosive Events in the Universe, Springer, p. 181

Trimble, V. Aschwanden, M.A. 2003. PASP 115, 514

Trimble, V. & McFadden, L. 1998. PASP110, 223

Trimble, V. Aschwanden M.A. & Hansen, C.J. 2007. Space Science Reviews 132, 1

Tsiganov, E.P. & Tsiganov, O.E. 2009. Nauka I Studia Vol. 2

Umemura, M. & Ikeuchi, S. 1986. Astrophysics & Space Science 119, 243

Valentijn, E.A. 1990. Nature 346, 152

van de Hulst, H.C. et al. 1957. BAN 14, 1

van den Bergh, S. 1960. MNRAS 121, 387 and ApJ 131, 558

van den Bergh, S. 2001 in V. Martinez et al. Eds. Historical Development of Modern Cosmology, ASP Conf. Ser. 252, p. 75

Vekeschagin, G.V. & Yegorian, G. 1998. International Journal of Modern Physics D17, 203

Wagoner, R.V. 1973. ApJ 179, 343

Wagoner, R.V. Fowler, W.A. & Hoyle, F. 1967. ApJ 148, 3

Walsh, D.R., Carswell, R. & Weymann, R. 1979. Nature 279, 38

Wheeler, J.A. 1968. American Scientist 56, 1

White, S.D.M & Rees, M.J. 1977. MN 183, 341

Zabrenowski, M. 1986. Astrophysics & Space Science 117, 179

Zeldovich, Y.B. 1968. Soviet Physics Uspekhi 11, 381

Zeldovich, Y.B. 1972. MNRAS 160 1p

Zeldovich, Y.B. & Novikov, I.D. 1966. A.Zh. 43, 750, Sov. Astr. AJ 10, 602

Zeldovich, Y.B. et al. 1974, JETP 40, 1

Zwicky, F. 1933. Helvetia Physica Acta 6, 110

Zwicky, F. 1937 ApJ 86, 217

Zwicky, F. 1951, PASP 63, 61