CSIRO PUBLISHING
Structure and Dynamics in the Local Universe
www.publish.csiro.au/journals/pasa
Publications of the Astronomical Society of Australia, 2004, **21**, 385–389

The Future of Astronomy and Physical Cosmology

P. James E. Peebles[A]

[A] Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, USA.
Email: pjep@princeton.edu

Received 2004 May 3, accepted 2004 September 27

Abstract: Astronomy and cosmology have a substantial observational and theoretical basis, but our standard model still depends on some working assumptions. I comment on the nature of the issues behind these assumptions, the guidance we might find from past resolutions of such issues, and the models we might consider for the future of research in this subject.

Keywords: history — philosophy — sociology of astronomy

1 What are we Trying To Do?

Research in astronomy and physical cosmology is healthy: we have a broadly successful empirical and theoretical basis for our subject and a rich set of observationally driven issues to study. An assessment of where research is likely to be headed in the longer term must be informed by an opinion of what we are trying to accomplish in physical science. I take it to be the discovery of how to encode growing empirical knowledge into an increasingly tight, compact, and internally consistent set of rules. We have no guarantee that the real world operates according to simple laws that we can discover in successive approximations, but physical science has made enormous progress by acting as if this were so.

The state of development of this reductionist program is very different in different branches of physical science, of course. Maybe the meaning of life, and what it means to be conscious, are to be found within in the known and established laws of physics, but it will be a long time before we know for sure. Biophysicists are learning how to deal with deeply complex systems; it's not surprising that we don't see them paying much attention to issues of fundamental physics. Particle physicists, on the other hand, have a standard model that gives an account of a vast amount of experimental results within a theory that can be written down in a page of equations. This is a spectacular success story. We have an example of this style of research in cosmology: general relativity in linear perturbation theory, with a simple prescription for initial conditions, gives a critical prediction for the relation between the large-scale distributions of matter and the thermal 3 K cosmic background radiation that is beautifully confirmed by the measurements (Bennett et al. 2003 and references therein). Analyses of galaxy formation, on the other hand, have more the flavour of biophysics; since galaxies are complicated it no surprise to see controversies not only about what the observations may be telling us about how the galaxies formed but also about what the standard and accepted cosmology actually predicts happens on the scales of galaxies. That is, the character of research in astronomy tends to be intermediate between what is happening in particle physics and biophysics. This is one guide to our future.

We have another guide from history. I outline in Section 2 a picture for the development of physical cosmology through the exploration of bold hypotheses. We have had to discard many; the surprising thing is that some have become part of an observationally well-established world picture. Our standard model still depends on bold hypotheses, however, mainly in the dark sector, as discussed in Section 3. I present in Section 4 a picture for the long-term future of the goal of reducing the hypotheses and establishing the physics of astronomy and cosmology.

2 What are the Lessons from Experience?

The history of the interplay between theory and practice in our subject teaches us that we should pay attention to elegant hypotheses, because they sometimes lead to aspects of reality, and that it is prudent also to bear in mind that nature is quite capable of forcing adjustments of our ideas of truth and beauty.

Consider the discovery of the expansion of the universe. The first analyses, by Friedmann (1922) and Lemaître (1927), assumed Einstein's general relativity theory and his argument that the universe is close to homogeneous in the large-scale average. There was no observational support for homogeneity: Einstein arrived at this picture by an interpretation of Mach's principle that we still don't understand. There was little empirical support for general relativity theory: the one precision test was the precession of the perihelion of Mercury, and Friedmann and Lemaître were considering an enormous extrapolation from the length scale of the Solar System. Lemaître did have another hint: Hubble's proposal of a linear relation between galaxy redshifts and distances. Lemaître's very

10.1071/AS04054 1323-3580/04/04385

reasonable assessment was that the evidence for the linear relation is schematic at best, however. (The discovery paper is Hubble 1929; Lemaître 1927 had the concept from a lecture by Hubble.) In short, the proposal that the universe is expanding from a very dense state was exceedingly speculative. Hubble (1936, 1937) recognised this; that is why he pioneered the cosmological tests. Bondi & Gold (1948) and Hoyle (1948) were properly skeptical and gave us an alternative, the steady-state cosmology. Its lasting elegance is seen reflected in the inflation scenario. Now, however, against what I would call extremely long odds, the evidence clearly shows that the universe has expanded in a close to homogeneous way from a much denser state, just as Lemaître (1931) visualised.

Other stories have different endings, of course. Consider Einstein's (1917) cosmological constant, Λ. When Einstein (1931) saw that his homogeneity concept fits Hubble's redshift–distance relation without the need for Λ, he declared that this term is inelegant and unnecessary. Pauli (1958) agreed, as did Landau & Lifshitz (1951). Zel'dovich (1968) taught us that a quantum vacuum energy density that looks the same to any freely moving observer, consistent with special relativity, would have the velocity-independent form of Λ. Already in the 1920s, however, Pauli recognised that the quantum zero-point energy of the electromagnetic field at laboratory wavelengths is absurdly large compared to what is wanted in cosmology (as described in Rugh & Zinkernagel 2002). The situation is if anything even worse now that we have to add the latent heats of the phase transitions of standard particle physics. By the mid-1980s the consensus was that a detectable value of Λ would be a theoretical abomination: the only logical and reasonable situation is that some sort of symmetry has forced the vacuum energy density to vanish, and that $\Lambda = 0$. There were continued arguments in favour of Λ, some, with McVittie (1956), as a part of a logically complete general relativity theory, and others, with Sandage (1961) and Gunn & Tinsley (1975), as a way to understand the constraints on the distance scale, stellar evolution ages, and curvature of the redshift–magnitude relation. Now, again against long odds, Λ has become part of the standard and accepted cosmology (Bennett et al. 2003 and references therein), and there is a substantial and growing literature on the theoretical physics of Λ pictured as dark energy, a new term in the stress-energy tensor. The change in the standard cosmological model, from Einstein–de Sitter (where Λ and space curvature are negligibly small) to ΛCDM, was sudden, but the weight of evidence has been building for quite some time. (A more detailed review of the historical developments is in Peebles & Ratra 2003.)

You will have heard on occasion the admonishment to trust our theories. Indeed, that advice taken in 1917 would have kept you spectacularly on track for general relativity theory and for Λ. You would not have done as well if you had clung to Einstein's seemingly self-evident presumption that the universe is static, of course, or if in the 1950s you had followed the advice of some of the deepest thinkers in physics, including Albert Einstein, Wolfgang Pauli, and Lev Landau, on the subject of Λ. If you had been one of the handful of people active in research in cosmology in 1930, you would have had the opportunity to decide whether your choice for a working model for large-scale structure would accept Einstein's argument from elegance for a homogeneous universe or Charlier's (1922) argument from the empirical evidence for a clustering hierarchy (what is now termed a fractal universe). I think in 1930 I would have made the wrong choice. In his work on the cosmological tests in the 1930s Hubble chose spatial uniformity as a working assumption, and he sought to test the velocity interpretation of the galaxy redshifts. Sandage (1961) accepted the velocity interpretation as a reasonable hypothesis and, among other things, sought to distinguish between the expanding steady-state and Friedmann–Lemaître cosmologies. If you had been active in research in 1985 you would have had the chance to choose between the clearly more elegant Einstein–de Sitter cosmological model and the empirical case for a low density universe. A big part of the history of the development of cosmology and astronomy has been the art of choosing useful working hypotheses.

3 What is the Present Situation?

The observational basis for this subject is much firmer now, but our world picture still depends on some major working hypotheses, mainly about the nature of the dark sector.

People have been debating the meaning of the Λ term, and its possible relation to the quantum vacuum energy density, for a long time. There are now two independent lines of evidence for the detection of this curious term, from the redshift–magnitude relation Sandage championed in the 1960s, applied now to supernovae (Tonry et al. 2003 and references therein), and from the measurements of the large-scale distributions of galaxies and the 3 K thermal background radiation (Bennett et al. 2003 and references therein). Both assume general relativity theory, applied on length scales some fifteen orders of magnitude larger than the precision tests, and the second depends also on the CDM model for structure formation. The consistency of the results is a strong argument for both assumptions. Good science demands more detailed checks, of course. The development of tests of gravity physics and the dark sector will be followed with particular interest, because there are reasons to wonder whether we have the full physics of dark matter and how general relativity theory is going to deal with the issue of the vacuum energy.

The litany of the quantum vacuum energy is worth recalling. We know that the quantum zero-point energy of matter in atoms and molecules and solids makes a real and measurable contribution to binding energies. We have experimental evidence that energy is equivalent to active gravitational mass. We know the electromagnetic field at laboratory wavelengths is well described by standard textbook quantum mechanics. It is entirely reasonable therefore to expect that the zero-point energy of the electromagnetic field is real. But as we noted in the last Section, the integrated zero-point energy of the electromagnetic field at laboratory wavelengths is an absurdly large active gravitational mass density. What has gone wrong? The anthropic argument is that the vacuum energy density summed over all zero-point energies, all the latent heats of all the phase transitions, and the energies of everything else that is close to homogeneous, are a function of position, and that we live where conditions allow life as we know it, a location close to where the total vacuum energy density happens to pass through zero (e.g. Pogosian, Vilenkin, & Tegmark 2004). The other line of thought is that there is a problem with the application of Einstein's field equation, perhaps on the right-hand side — the prescription for the stress-energy tensor — or maybe on the left-hand side — the prescription for how spacetime responds to the contents of the universe. Whatever the fix, it will be a revolutionary change in our world view. The successes of the cosmological tests suggest that the fix will not greatly upset the accepted physics of extragalactic astronomy. This is a working hypothesis, of course.

Another hypothesis in the standard model for cosmology is that the dark matter acts like a gas of nearly free particles. Is the dark sector really this simple?

In the standard cosmology the energy density associated with Λ, or dark energy, is about 75% of the total, and dark matter contributes about 20%. The rest is in the visible sector: baryons amount to about 4% of the total, the relict thermal neutrino mass density is about 3% of the baryon density, and the energy in the relict 3 K thermal electromagnetic radiation is about 3% of the neutrino density. The baryons and the electromagnetic field, with a little help from neutrinos and gravity, including that of the dark sector, gave us galaxies, stars, planets, and people. The standard model for the dark sector has none of this complexity: dark matter just piles up in nearly smooth halos while the dark energy density remains close to homogeneous and maybe slowly evolves. Is this the way it really is, or only a rough approximation that we get away with because the gravity by which we explore the dark sector is a blunt probe? If the standard model for the dark sector is too simple we may discover it through anomalies in the theory and observation of structure formation. Two examples illustrate the situation.

In numerical simulations of the ΛCDM cosmology the considerable merging of dark matter halos at low redshifts can be taken to mean that the massive galaxies formed relatively recently. The observational implications of late formation of the giant galaxies are debatable; my objections are detailed in Peebles (2002). Analyses of structure formation in ΛCDM argue both ways. Baugh et al. (1998, 2003) find that in this cosmology the comoving number density of galaxies with present stellar mass $M_{\rm star} \gtrsim 10^{10.7}\,M_\odot$, about that of the Milky Way galaxy, has increased by about three orders of magnitude since redshift $z = 2$. In the analysis of Loeb & Peebles (2003) and Gao et al. (2004) the mass concentrations characteristic of the luminous parts of the giant galaxies were assembled at $z \sim 6$, and the stellar masses were about half the present values at $z \sim 3$. If the latter is the correct story it removes this issue from my list of apparent anomalies. It will be interesting to see how the community resolves these very different interpretations of a complex process, and what the observations will teach us about the vigorous ongoing merging at $z < 1$ that all agree is predicted by the standard cosmology.

My second example is the prediction in the ΛCDM cosmology that there are substantial numbers of low mass dark matter halos in groups of normal L_* galaxies and in the voids defined by the normal galaxies. The latter is the basis for the original biasing picture for the CDM model in an Einstein–de Sitter cosmology: the bulk of the mass would be in debris between the concentrations of normal galaxies and so would not affect the small-scale relative velocities of normal galaxies (Davis et al. 1985). Tully (2005) emphasises that the situation in mass concentrations is complicated — the mass-to-light ratio depends on the crossing time — but the issue here is what is in the voids. Dekel & Silk (1986) point out that the debris in voids might reasonably be expected to be observable as dwarf or irregular galaxies. Blitz et al. (1999) point out that the debris might also be expected to be observable as HI clouds, and they suggest that the high velocity clouds may in part be this debris, falling onto the Local Group from distances ~1 Mpc. In the new standard low mass density cosmology the separation of mass and light has to be considerably less strong, but the theoretical argument for debris in the voids still seems persuasive. On the observational side, there are HI and HII regions at distances ~30 kpc from the luminous parts of galaxies (Ryan-Weber et al. 2004), but these plausibly are parts of the galaxy rather than infalling debris. The object DDO 154 is a galaxy, but with a large ratio of HI mass to stellar luminosity. It might be considered a prototype 'dark galaxy' (Carignan & Freeman 1988; Carignan & Purton 1998), of the type one would be inclined to look for in the voids. The examples of dark galaxies I know are in concentrations of normal galaxies, however, not in voids. Observers have been emphasising for a long time that dark galaxies are striking rare (Briggs 2003; Pisano et al. 2003; de Blok et al. 2002; and references therein), contrary to what seems to be a reasonable interpretation of the theory.

The discrepancies between the theory and observation of irregular and dark galaxies may be only apparent, a result of the difficulty of analysing complex processes, or it may be a clue to better physics in the dark sector. There is no shortage of ideas about new physics: examples are warm dark matter (Bode, Ostriker, & Turok 2001), possibly supplemented by small-scale primeval isocurvature departures from homogeneity (Sommer-Larsen et al. 2004), and self-interacting dark matter, perhaps with a short range of interaction (Spergel & Steinhardt 2000), or perhaps a long-range interaction that acts as a fifth force in the dark sector. (The history of this last line of thought is reviewed in Farrar & Peebles 2004.) We will choose the working hypotheses, and future generations will judge the results, maybe in the style of Section 2.

4 How Will we Know When we are Done?

Advances in the astronomy of cosmic structure formation and evolution, and in the understanding of how the observations constrain the theory, may support the hypotheses of the present standard model or drive adjustments. Either way, astronomy and cosmology likely will converge to completeness, though in a somewhat different sense from what people were considering at the end of the 19th century (Badash 1972). There is no way to know that science has reached some ultimate theory that underlies all theories: our methods can at best establish empirical evidence for successive approximations. There could come a time when the theory, and the methods of application of the theory, are good enough for any measurement that it is practical to make. That would complete that branch of research. Studies of the very early universe could arrive a theory that is perfectly internally consistent and agrees with all the constraints from particle physics and astronomy, by the clever choice of sufficient numbers of theoretical constructs, leaving no critical test of the traditional kind. We would not know whether this theory describes what really happened or is just a good fitting function. The prospect is dismal, but one people have faced ever since they started doing science: no one ever promised us that the world is comprehensible. Hints kept turning up, leading to deeper science.

The end game for astronomy will not follow the dismal mode. Consider that in standard and exceedingly well tested physics all protons are alike. That is why you do not hear about richly detailed studies of the provenance of protons captured as cosmic rays. Tables of critical astronomical data could be taken to read that all galaxies with the same morphology and luminosity are alike. This is a successful example of the art of extracting simple and instructive regularities out of complex systems. But each galaxy has a memory of the conditions under which it formed, and of the contingencies of its evolution (a term I learned from Stephen Jay Gould's instructions of how to think about something really complex, the processes of the origin and evolution of life). We are hearing about realistic plans to supplement global statistics such as galaxy luminosity functions with the rich details that are becoming available (Freeman & Bland-Hawthorne 2002; Driver 2004). This includes the environments of individual galaxies; the formation of star populations, gas, dust, black holes, and cosmic rays out of the initial conditions of individual galaxies; the census of nearby planets; the identification of those capable of supporting life like us or maybe very different; and surveys of the forms of life that have chosen to communicate by radio broadcast. The list of things to do is in practice unlimited: astronomy will end in myriad details.

Brent Tully is flourishing in a golden age for astronomy: the subject was a lot less exciting a century ago and likely will become pretty dull by the time people have mapped out in detail the signatures of formation of a fair number of nearby galaxies. But I think it is clear from the holes in our present standard model that the golden age is going to last for quite a long time to come.

Acknowledgments

This paper grew out of remarks at a workshop on the occasion of Brent Tully's birthday. I am indebted to him, the organiser Joss Bland-Hawthorn, and the participants for the occasion and inspiration for this essay.

References

Badash, L. 1972, Isis, 63, 48
Baugh, C. M., Benson, A. J., Cole, S., Frenk, C. S., & Lacey, C. 2003, in The Mass of Galaxies at Low and High Redshift, Venice, 24–26 October 2001, p. 91
Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998, ApJ, 498, 504
Bennett, C. L., et al. 2003, ApJS, 148, 1
Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, D., & Burton, W. B. 1999, ApJ, 514, 818
Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93
Bondi, H., & Gold, T. 1948, MNRAS, 108, 252
Briggs, F. H. 2003, in IAU Symposium 217, Recycling Intergalactic and Interstellar Matter, Sydney, 14–17 July 2003
Carignan, C., & Freeman, K. C. 1988, ApJ, 332, L33
Carignan, C., & Purton, C. 1998, ApJ, 506, 125
Charlier, C. V. L. 1922, ArMAF, 16, 1
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
de Blok, W. J. G., Zwaan, M. A., Dijkstra, M., Briggs, F. H., & Freeman, K. C. 2002, AA, 382, 43
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Driver, S. 2004, PASA, 21, 344
Einstein, A. 1917, Sitzungsber Preuss Akad Wiss, p. 142 [English translation in The Principle of Relativity (Dover, New York, 1952), p. 177]
Einstein, A. 1931, Sitzungsber Preuss Akad Wiss, p. 235
Farrar, G. R., & Peebles, P. J. E. 2004, ApJ, 604, 1
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Friedmann, A. 1922, ZPhy, 10, 377
Gao, L., Loeb, A., Peebles, P. J. E., White, S. D. M., & Jenkins, A. 2004, ApJ, 614, 17
Gunn, J. E., & Tinsley, B. M. 1975, Natur, 257, 454
Hoyle, F. 1948, MNRAS, 108, 372
Hubble, E. P. 1929, PNAS, 15, 168
Hubble, E. P. 1936, The Realm of the Nebulae (New Haven, Yale University Press)
Hubble, E. P. 1937, MNRAS, 97, 506
Landau, L. D., & Lifshitz, E. M. 1951, The Classical Theory of Fields (Pergamon, Oxford), p. 338
Lemaître, G. 1927, ASSB, 47A, 49
Lemaître, G. 1931, Natur, 127, 706
Loeb, A., & Peebles, P. J. E. 2003, ApJ, 589, 29
McVittie, G. C. 1956, General Relativity and Cosmology (Chapman & Hall, London), p. 35
Pauli, W. 1958, Theory of Relativity (Pergamon, New York) [reprinted 1981 (Dover, New York), p. 220]
Peebles, P. J. E. 2002, ASP Conf. Ser. 283: A New Era in Cosmology, 351
Peebles, P. J. E., & Ratra, B. 2003, RvMP, 75, 559

Pisano, D. J., Barnes, D. G., Gibson, B. K., Staveley-Smith, L., & Freeman, K. C. 2003, in IAU Symposium 217, Recycling Intergalactic and Interstellar Matter, Sydney, 14–17 July 2003
Pogosian, L., Vilenkin, A., & Tegmark, M. 2004, JCAP, 7, 5
Rugh, S. E., & Zinkernagel, H. 2002, StHPM, 33, 663
Ryan-Weber, E. V., Meurer, G. R., Freeman, K. C., Putman, M. E., & Webster, R. L., et al. 2004, AJ, 127, 1431
Sandage, A. 1961, ApJ, 133, 355
Sommer-Larsen, J., Naselsky, P., Novikov, I., & Gotz, M. 2004, MNRAS, 352, 299
Spergel, D. N., & Steinhardt, P. J. 2000, PhRvL, 84, 3760
Tonry, J. L., et al. 2003, ApJ, 594, 1
Tully, R. B. 2005, ApJ, in press (astro-ph/0312441)
Zel'dovich, Ya. B. 1968, UsFiN, 95, 209 [English translation in 1986, SvPhU, 11, 381]