CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-15526-8 — Deconstructing Cosmology
Robert H. Sanders
Excerpt
More Information

Introduction

Under the sub-subject of cosmology, Amazon.com currently lists 5765 items. Among them there are textbooks, serious scientific discussions, popular books, books on history, philosophy, metaphysics and pseudoscience, mega-bestsellers like those by Stephen Hawking, Brian Greene and Lisa Randall, and works no one has heard of by authors as obscure as their books. It would almost seem as though the number of books on the subject is expanding faster than the Universe; that soon the nature of the missing mass will be no mystery – the dark matter is in the form of published but largely unread cosmology books. Does the world need yet another book about this subject? Why have I decided to contribute to this obvious glut on the book market? Why do I feel that I have something to add of unique value?

The idea for the current project had its dim origins in the year 2003 when I was invited to lecture on observational cosmology at a summer school on the Aegean island of Syros. I was surprised at this invitation because I am neither an observer nor a cosmologist; I have always worked on smaller-scale astrophysical problems that I considered soluble. In this career choice I was no doubt influenced by my first teachers in astrophysics, who were excellent but traditional and, to my perception at least, found cosmology to be rather fanciful and speculative (although I never heard them explicitly say so and almost certainly they would not say so now).

But I decided that this invitation was an opportunity to learn something new, so I prepared a talk on the standard cosmological tests (e.g., the Hubble diagram, the angular size–redshift relation and the number counts of faint galaxies) in the context of the current cosmological paradigm that is supported by modern observations, such as the very detailed views of tiny anisotropies in the cosmic microwave background radiation (CMB). The issue I considered was the overall consistency of these classical tests with the standard model – Lambda-CDM (ΛCDM).

I was actually more interested in finding inconsistency rather than consistency. This is because of my somewhat rebellious nature, as well as my conviction that science primarily proceeds through contradiction and conflict rather than through agreement and "concordance." However, somewhat to my chagrin, I found that the classical tests were entirely consistent with the standard paradigm, although with much lower precision than that of the modern CMB observations. There is no conspicuous inconsistency between the model and those observations considered to be cosmological.

And yet worries persist (not only by me) about Lambda-CDM, not because of any direct observational contradiction on a cosmological scale but because of the unknown nature of the two dominant constituents of the world – dark energy and dark matter. These two media are deemed necessary because the Universe, and more significantly, objects within it, do not behave as expected if the laws of physics on the grand scale take on the same form as that established locally. The universality of terrestrial and Solar System physical law deduced from local phenomena is an assumption, and one that is difficult to test. The required existence of these two invisible media – aethers, undetectable apart from their dynamical or gravitational effects – is perhaps a hint that the assumption of the universal validity of local physics may not be valid.

These two components – dark energy, designated by Λ, and dark matter, abbreviated as "CDM" (for cold dark matter) – comprise 95% of the energy density of the Universe, and yet we have no clear idea of what they are. Dark energy is elusive; the only evidence is, and very possibly can only be, astronomical. Is this simply a cosmological constant in Einstein's equations, or is it a vacuum energy density – the zero-point energy of a quantum field? Is it the evolving energy density of a background field – a field that is not usually included in the formulation of general relativity? Or does apparent requirement of this medium signal the breakdown of general relativity itself on a cosmological scale?

Dark matter appears to be required to explain the details of the observed temperature fluctuations in the cosmic microwave background radiation and, more fundamentally, the formation of structure in an expanding universe of finite lifetime. This same dark matter presumably clusters on the scale of galaxies and accounts for the total dynamical mass of these systems – a mass that often exceeds the detectable baryonic mass by a large factor. There is the additional problem that the dark matter particles, which should be detectable locally, have so far not made an appearance in various increasingly sensitive experiments designed to catch them, but this, I have argued, is not (and cannot be) a falsification.[1]

For the two components taken together, there is a further complication of naturalness: the density of these two "fluids" decreases differently as the Universe expands (they have different equations of state). The dark matter dilutes with the expanding volume element; the dark energy does not dilute at all (or at least very differently) with expansion. Why should these two components with different equations of state have, at the present epoch, after expansion by many factors of ten, energy densities that are comparable? This may well indicate that the models forced upon the observations are inappropriate.

There are reasonable arguments that the "mysteries" surrounding these two unknown constituents of the Universe are not really so profound and do not indicate a failure of the cosmological paradigm or of general relativity. The postulates of Einstein's general theory do permit a cosmological constant, and its phenomenological impact is the observed accelerated expansion of the Universe. The small magnitude of the constant with respect to the expectations of quantum physics does not reflect an incompleteness of general relativity but rather a lack of understanding of the nature of quantum fields in curved space. The near coincidence of the densities of matter and dark energy is not so surprising and will occur over a large range of cosmological time.[2] The dark matter particles will eventually be discovered, revealing their identity and nature.

In my opinion, the essential problem with the paradigm is that cosmology, via dark matter, impinges directly upon the dynamics of well-studied local systems – galaxies – and here, I will argue, the cosmological paradigm fails. The most profound problem, specifically with CDM, is the existence of a simple algorithm, with one new physical constant having units of acceleration, that allows the distribution of force in an astronomical object to be predicted from the observed distribution of observable matter – of baryons. These predictions are surprisingly accurate for spiral galaxies, as evidenced by the matching of observed rotation curves where even details are reproduced. This demonstrates that the total force precisely reflects the distribution of observable matter even in the presence of a large discrepancy between the traditional (Newtonian) dynamical mass and the observable mass. In the context of dark matter this fact would seem to imply a very close interaction between the dark and detectable matter, an interaction totally at odds with the nature of dark matter as it is perceived to be – a dissipationless fluid that interacts with normal matter primarily by gravity.

That algorithm, of course, is modified Newtonian dynamics (MOND), proposed more than thirty years ago by Mordehai Milgrom, and its success constitutes the greatest challenge for dark matter that clusters on the scale of galaxies.[3] In so far as such dark matter is an essential aspect of ΛCDM, MOND directly challenges the prevailing paradigm, and that becomes more than a pure scientific problem; it is a significant sociological issue. This is because ΛCDM has become something of an official religion, outside of which there is no salvation and beyond which there is only damnation. The entire phenomenology of galaxies that is described so well by MOND is systematically dismissed or relegated to the category of problems for the future. This reveals a danger to the creative process presented by a dogma that is too widely and too deeply accepted. However, here I wish to discuss not only this danger, but more positive possibilities for physics and cosmology, revealed by the perceived successes of both the ΛCDM and MOND paradigms. This, to answer the question that I posed above, is the reason that I began this project.

First of all, I would like to state clearly what this book is not. It is not a critique of modern science or the application of the scientific method to the study of cosmology. At present, there are many such absurd criticisms of the methods and results of science – so much so that one can identify an anti-rationalist or even anti-realist sentiment that is dangerous in modern society. I do not wish to be identified with those who deny many of the important results of modern physics as these relate to cosmology – from relativity to particle physics to the overall coherence of the current cosmological scenario.

The title of the book might suggest to some that from the onset I have a negative attitude toward the study of cosmology. This is not what I wish to imply by use of the term "deconstructing." Following the second definition given by the Merriam-Webster dictionary, to deconstruct is to take something apart, in this case the standard paradigm, in order to reveal possible flaws, biases or inconsistencies, and that is what I wish to achieve. Overall, I believe the reader will find that I am positive about the application of the scientific method to the study of the entire Universe; the achievements of humans in the past century in understanding the structure of the world on the grand scale are truly remarkable. But there remain, at the very least, empirical problems for the observed structure and kinematics of bound astronomical objects, primarily galaxies. I will criticize these aspects of the standard cosmological paradigm as well as the general sentiment that there is no room left for doubt.

I begin with a comparison of the standard Big Bang model to early human creation mythologies, emphasizing those aspects that are similar and those that are different. This leads inevitably to a general discussion of the philosophical issues raised by cosmology as a science and the ways in which the subject differs from the other physical sciences. After this introduction, there follows a historical interlude on the origins of Big Bang cosmology, its predictive basis and its remarkable success on an empirical level. Included here is a chapter on the proposed phenomenon of "inflation" in the very early Universe – the motivation of the idea as well as its elevation to an essential aspect of the cosmological paradigm in spite of the absence of a clear microphysical basis. Going beyond the essential Big Bang, I discuss "precision" cosmology and the "concordance" model – ΛCDM – emphasizing its consistency in explaining the observed anisotropies in the cosmic microwave background, the large-scale distribution of galaxies and the accelerated expansion of the Universe.

Finally I deal with what I see as the principal problem of the paradigm: the mysterious nature of the two dominant components – dark energy and dark matter.

In particular I dwell upon the failures of cold dark matter in explaining the basic photometric and kinematic observations of galaxies. I discuss at length the leading alternative to dark matter – modified Newtonian dynamics (MOND), its observational success and the challenge to the cosmological paradigm presented by this simple algorithm. I consider the possible modifications of physics required by MOND as well as the perceived problems of MOND with respect to cosmology. I discuss possible reconciliation of MOND on galaxy scales with the evidence for dark matter on a cosmological scale.

In conclusion (and here and there throughout) I delve into philosophical issues. But I am not a philosopher and have never studied philosophy (apart from a survey course in my second year of college – long ago). So for me, any discussion of philosophy is to enter the realm of dangerous speculation. In general this is a risky business for a scientist. On the other hand, it is difficult to avoid philosophy in any discussion of cosmology.

I have tried to omit serious mathematics throughout – essentially no formulae – but I do use standard scientific notation for the very large and small numbers encountered in astronomy and cosmology: 1.38×10^{10} years for the age of the Universe (instead of 13 800 000 000) or 10^{-32} seconds for the duration of the early inflationary period (instead of 0. [followed by 31 zeros] 1).

I also describe the mass of sub-atomic particles in energy units (via Einstein's famous formula $E = Mc^2$). As is usual, the preferred unit is the electron volt or eV ($= 1.6 \times 10^{-12}$ ergs). And of course there are keV (kilo-electron volts or one thousand eV), MeV (mega-electron volts or one million eV) and GeV (giga-electron volts or one billion eV).

For units of distance I typically use light years – the distance light travels in one year, or about 10^{18} centimeters. And of course there are kilo-light years for galactic scale distances and mega-light years for cosmic scales. In places – typically in published figures – I revert to the more astronomical distance units of parsecs (about three light years). And then there are kiloparsecs (kpc) and megaparsecs (Mpc).

Finally, a word on terminology. I use the word "world" generally to be synonymous with Universe or Cosmos. But when discussing the "multiverse," world will refer to not just our piece of it, but to the whole "shebang."

1

Creation Mythology

> When the sky above was not named,
> And the earth beneath did not yet bear a name,
> And the primeval Apsû, who begat them,
> And chaos, Tiamat, the mother of them both,
> Their waters were mingled together,
> And no field was formed, no marsh was to be seen;
> When of the gods none had yet been called into being.

So begins the Enuma Elish, the seven tablets of creation, describing the ancient Sumerian creation myth.[1] In the beginning the world is without form, and fresh water (Apsû) and salt water (Tiamat) mingle together. Then, in an act of creation, there follow six generations of gods, each associated with a natural manifestation of the world, such as sky or earth. Light and darkness are separated before the creation of luminous objects: the Sun, the Moon, the stars. The sixth-generation god, Marduk, establishes his precedence over all others by killing Tiamat and dividing her body into two parts – the earth below, and the sky above. He establishes law and order – control over the movement of the stars, twelve constellations through which the Sun and the planets move – and he creates humans from mud mixed with the blood of Tiamat.

The similarity to the Hebrew creation mythology described in the book of Genesis has long been recognized:[2] In the biblical story creation takes place in six days, corresponding to the six generations of "phenomenon" gods in Babylon, and the separation of light and dark precedes the creation of heavenly bodies. There is an initial homogeneous state in which the various constituents of the world are mixed evenly together, and an act of creation at a definite point in time – an act which separates these constituents and makes the world habitable (and more interesting).

These aspects are also evident in the Greek creation mythology[3] in which elements of the world are initially mixed together in a formless way – Chaos.

However, at some point, two children are born of Chaos – Night and Erebus, an "unfathomable depth where death dwells" (not an obvious improvement over the initial state of Chaos). But then, also in an unexplained way, something positive and truly magnificent happens: Love is born and Order and Beauty appear. Love creates Light and Day and the stage is set for the creation of the Earth and the starry heavens. Again, separation of light and darkness occurs before the creation of earthly and heavenly objects.

These ancient myths appear to presage the modern scientific narrative of creation – the Big Bang. Here, at a definite point in time, 13.8 billion years ago, the creation of the world occurs. There is no specified prior state and no creating entity; the event is apparently spontaneous.[4] Initially the world is extremely dense and hot but rapidly expands and cools. An instant after the creation event the Universe is recreated in an exponential expansion driven by the energy density of the vacuum; this short-lived inflationary epoch drives the space–time to be very nearly smooth and topologically flat, but at the same time creates tiny fluctuations that eventually will appear as the observed structure of the Universe. A few seconds after creation, when the temperature has fallen to a few tens of billions of degrees, the Universe primarily consists of radiation (photons), electrons and positrons (leptons), ghostly, almost massless neutrinos, a small trace of protons and neutrons (baryons) that will become the primary observable component of the world, and more massive particles that interact very weakly with other components. These massive particles are effectively dark and will forever remain so, although they will come to dominate the dynamics of the matter and the structure that will form from it. After two or three minutes, the light elements (deuterium, helium, lithium) are synthesized via nuclear fusion of about 11% of the protons (plus the remaining neutrons), and the electrons and positrons (anti-electrons) annihilate, leaving a small residue of electrons. After several hundred thousand years, when the Universe has cooled to a few thousand degrees, the protons combine with the electrons, becoming neutral hydrogen, and the radiation (light) decouples from the matter; one could say that light separates from darkness. The structure that is currently observable in the Universe, stars and galaxies and clusters of galaxies, is then free to form via gravitational collapse.

In interesting and perhaps meaningful respects, the form and language of the modern cosmological paradigm is similar to these ancient creation mythologies. In the examples above, there is a time sequence which monotonically and irreversibly progresses from past to present, and there is a definite point in the past at which creation occurs; the world has a beginning. Further, there is initially a more homogeneous state in which the various components are mixed together uniformly; there follows a separation of these components and a separation of light and darkness predating the appearance of astronomical objects; and finally there is the formation of the structures of the present recognizable world. This similarity in the ordering of events – the linguistic framework upon which the creation stories are placed – probably reflects the fact that both the ancient creation myths and the modern cosmological paradigm are products of human thought processes, and limitations of experience and language restrict the description of cosmological phenomena. It would appear as though the language of the scientific paradigm has been filtered through the narrative established by these particularly occidental creation mythologies in ways that are not intentional or conscious. After all, how does one conceive of the spectacular occurrence of the Big Bang, other than through the framework of stories that are in our direct tradition?

But in this connection one should keep in mind that neither in creation mythologies nor in scientific scenarios is it always the case that there is a single universe with a definite beginning. We can identify two trends that run counter to one another: the concept of a temporally finite single world and that of a world that is static eternal and possibly multiple – a universe of universes – co-existing or cyclic. In Hindu creation mythology, the world is created and destroyed and re-created in cycles corresponding to a day and night in the life of Brahma – at 8.6 billion years, amazingly close to the cosmological timescale of the Big Bang.[5] These many sequential universes are rather closely mimicked by current ideas such as eternal inflation and the multiverse scenario, but here the universes are not so much sequential as co-existing (in some extended sense of simultaneity) in different parts of a possibly higher-dimensional infinite stage. But, in the property of its seething, percolating nature, such a world is constant and unchanging. The idea of a universe that is eternally immutable in its general aspect has had great appeal from Newton to Einstein to Hoyle to Linde.

The creation scenarios, ancient and modern, reflect the human drive and desire for an understanding of the Universe, its origin and evolution, and the place of human consciousness within it. However, given the predictive success of modern science and the retreat of superstitious or theistic explanations for the unknown, physical cosmology has achieved a new pre-eminence and dominance in providing a true picture of the Universe – its formation and evolution. With the development of new tools for observing the Universe as a whole, cosmology has become a proper and respectable science. This is particularly true since the discovery of the cosmic microwave background radiation, the CMB – that faint glow of an earlier, hot universe.

In the early twentieth century the development of general relativity, Einstein's theory of gravity, made it possible to "do cosmology" with physics. The new observational tools have now made it possible to "do physics" with cosmology – or, at least, that is the perception. But to what extent do we understand the constituents and evolution of the Universe, and to what extent can we draw conclusions on physics from the cosmology of one universe? Is cosmology a proper physical science?

Certainly in one essential respect cosmology differs from other more traditional sciences: the study of the Universe as a whole lends itself to considerations that are more generally taken to be philosophical rather than purely scientific or narrowly empirical. First of all, there is the issue of reductionism. There is an essentially reductionist predilection in contemporary science in general (a preconception that is not fundamental to the scientific method). Of course, there is a process of unification that is basic to science – a search for efficiency in explanation. But this has been taken quite far: all follows from a few basic laws or a "theory of everything," and to derive such a theory has become a modern holy grail. Even biological evolution and the development of consciousness can be calculated from a few first physical principles (although in practice, this may be impossibly difficult). The reductionist predilection elevates cosmology to a primary role, with large-scale phenomena taking precedence over the phenomenology of individual objects such as stars or galaxies. This prioritization minimizes the possible role and importance of emergent phenomena – the idea that many constructs considered to be basic – even space and time – may emerge, in a way that is not obvious, from more fundamental components. That is to say (with Aristotle) that the total may be more than the sum of its parts.[6]

Emergence may be divided into two categories: "weak" and "strong". Weak emergence is that which can be computed; i.e., the emergent phenomenon can be calculated from the behavior of the primary constituents; weak emergence is not necessarily at odds with reductionism. On the other hand, strong emergence cannot, in principle, be computed and therefore is the more metaphysically challenging concept. But this distinction between strong and weak is in fact blurred by the practical difficulty (or impossibility) of computing complex phenomena from the laws governing the behavior of more basic components. Is it possible to compute the subsequent development of the Universe in all of its present complexity from a snapshot recorded in the cosmic microwave background at the single epoch of decoupling of matter and radiation? In cosmology, the very possibility of emergent phenomena challenges reductionism and the assignment of priority to cosmology over the study of the more complex constituents of the Universe.

A second issue is that of purpose or teleology: modern science is essentially naturalistic and non-teleological, which is to say that evolution, cosmological or biological, has no purpose or final goal but proceeds by natural law without intervention from supernatural or theistic agents. This issue, that of teleology, is a metaphysical (and political) minefield that is best avoided by the practicing scientist. Indeed, it is not possible to be a functioning scientist without being, at least in a methodological sense, a naturalist. But the idea of a purposeless universe would have appeared inconceivable to classical philosophers such as Plato and Aristotle; it is essentially a modern assumption that cannot be proven in a purely scientific context. The possibility of a purposeful (but not necessarily theistic) universe is related to the issue of fine-tuning of fundamental physical constants and cosmological parameters: Why is it that the Universe is conducive to the development of life forms capable of understanding it? A teleological explanation would have profound implications for cosmological and biological evolution, but it is difficult to conceive of how such a concept could be operationally included within the scientific method.

A third issue is that of cosmological space and time. It can be said that the theory of relativity (special and general) destroyed the Newtonian concept of absolute space and time. However, in some sense, physical cosmology has restored it. There is, at least operationally, a preferred universal frame – that which is at rest with respect to the cosmological background – and there is a preferred cosmic time with a preferred direction that apparently proceeds smoothly and irreversibly from the beginning to the present. Time is particularly problematic: Does cosmic time correspond to other measures of time, atomic or biological? Is the direction of time strictly a matter of the entropy of a complex system or does some *a priori* structure (or broken symmetry) set this irreversibility?

Finally, with respect to "doing physics" with cosmology, there is the one-universe problem – the fact that usual statistical arguments cannot be used when confronted with a sample size of one object. Of course, in discussing constituents such as density fluctuations, their formation and evolution, statistical methods are appropriate as long as these fluctuations have a size that is small compared to the limit of the observable Universe – as long as there are many such fluctuations. Then observations of fluctuations and their interpretation belong more properly to the realm of astronomy. However, when the observable fluctuations approach the horizon, the size of the observable Universe, then there are only several in the sample. This is the well-known problem of "cosmic variance," but here again the issue is more fundamental: as in teleology there is a question of "naturalness" or "fine-tuning." Why is the Universe tuned to produce intelligent life?

This question lends itself to an anthropic interpretation. That is to say, the Universe must have quite precisely the properties that it does have (initial conditions, values of fundamental constants, number of dimensions) because these are the very properties that allow cogent observers (us) to develop and pose such questions.[7] There has been much discussion of this idea over the past four decades; supporters argue that it is profound while opponents assert that it is trivial. While the anthropic principle may in some sense be true, it does have the effect of choking off further research into naturalistic explanations for the world being as