IMAGE AND REALITY IN COSMOLOGY

Simon D.M. White
Max-Planck-Institut für Astrophysik
Garching, Germany

Because of the nature of the objects they treat, astronomers and cosmologists build and test theoretical models in a different way from scientists in other domains. In the following essay I discuss how these differences blur the distinctions between image and reality and observer and observed.

1) The limitations of astronomy and cosmology

Astronomy is distinguished from the other sciences by the enormous sizes and distances of the objects it studies and the very long timescales on which they evolve. Astronomers are unable to manipulate stars or galaxies, and so to prepare and carry out experiments of the kind which are the foundation of most physical sciences. The role of "experiment" has to be replaced by "observation". Celestial objects are made visible to us by many forms of radiation — X-rays, normal light, radio waves, even neutrino emission and perhaps, one day, gravitational waves. This radiation must be collected, analysed, classified and somehow organized into a coherent physical picture of distant and unapproachable systems. The resulting picture is always a simplified model, a caricature which encapsulates only those aspects of the objects which predominantly affect the particular light in which they are viewed.

A few other sciences and social sciences have similarly limited experimental capability and so are forced to proceed primarily through observation. In some of these, for example, anthropology or economics, the natural timescales are short enough that it is possible to observe directly the interplay of cause and effect. In evolutionary biology, geophysics or astronomy, however, the timescales of interest are usually far longer than a human lifespan. The nature of evolution can then only be inferred indirectly through examination of the fossil record. The finite speed of light allows astronomers direct access to a kind of fossil record. Distant objects are seen not as they are today but as they were when their light left them. For the most distant galaxies currently known, the light travel time is more than 80% of the age of the Universe, and they are therefore seen when less than one fifth the age of nearby galaxies.

Building a physical model on the basis of observations of this kind requires astronomers to adopt simplifying principles. The first and most important of these is that the laws of physics are the same everywhere. This is, in part, a definition of what is meant by the laws of physics, but it is also usually given the more restrictive meaning that all the constants of physics, for example, Newton's and Planck's constants and the masses and changes of the elementary particles, are independent of time and place. Astronomy actually provides abundant checks of this last hypothesis. Thus spectra of distant galaxies reveal that they are made of the same chemical elements as the Earth and that the atomic structure of these elements is identical to very high precision.

A second simplifying assumption is needed before observations of distant "young" galaxies can be used as a fossil record on which to base evolutionary models. The population of galaxies seen at a given distance must be assumed statistically equivalent to the ancestor population of nearby galaxies at the corresponding historical epoch. There is, of course, no sense in which distant glaxies can be considered the actual ancestors of nearby systems. The situation is thus very different from that of the geological fossil record where some of the bones found in low-lying strata do very plausibly belong to the ancestors of present-day creatures. This assumption that statistically equivalent galaxy populations evolve in different places is in fact an assumption about the overall homogeneity of the Universe on large scales.

Cosmology is the scientific study of the structure and evolution of our Universe as a whole. Because, by definition, we can only observe one Universe, cosmologists do not have the luxury of comparing universes with different evolutionary histories in order to isolate the underlying physical processes which shape them. Furthermore information is not immediately available for all places and epochs in the Universe but only for those which lie along our past light core and so are directly observable. Even stronger simplifying assumptions are therefore needed to organise the available information into a simple and testable model. In this context the assumption that the Universe in the large looks the same at all places and in all directions has been enshrined as the "Cosmological Principle", an extension of the Copernican argument that our location in the Universe is not privileged to say that all locations, including our own, should look the same. Again some (but not all) aspects of this assumption can be checked by direct observation. Galaxy populations at large distances but in different directions do indeed appear very similar and to have had very similar histories.

2) The Standard Model of Cosmology

As a result of the limitations I have just described, the current standard model of cosmology is a simple one. Three aspects of it are relevant for my purposes in this lecture: its geometry, its physical content, and the origin and nature of the small-scale structure within it. I now address each of these briefly in turn.

The Cosmological Principle states that the Universe should look the same from all positions and in all directions. Within General Relativity, Einstein's theory of gravity, the exact enforcement of this requirement leads to a very restricted set of world-models, the so-called Friedman-Robertson-Walker models. The spatial geometry of these models can either be positively curved (like the surface of a sphere), flat (like Euclid's space) or negatively curved (like a saddle). In the first case the volume of the Universe is finite, in the latter two it can be infinite. All these models can change size with time, but their global geometry remains fixed. The size evolution is determined by the material content of the Universe and how it is affected by the cosmic expansion. Different hypotheses lead to universes which exist eternally in the future, in the past, or in both.

Even at this very simplified level it is possible to gain fundamental insights into cosmic evolution as is shown by the demise of the Perfect Cosmological Principle. This extension of the Cosmological Principle states that the Universe should look similar at all locations, in all directions and <u>at all times</u>, and it led to the formulation of the Steady State Theory in the 1940's and 50's. Observations of very distant galaxy populations show them to differ dramatically from galaxies seen nearby, thereby demonstrating that the properties of our Universe do change with time and killing the Steady State Theory.

A number of different components appear to contribute significantly to the current content of the Universe. Ordinary matter, consisting of protons, neutrons and electrons, makes up all the planets, stars, galaxies and intergalactic material that we are able to observe directly. Electromagnetic radiation, detected directly as the Cosmic Microwave Background (CMB), currently contains only a fraction of a percent of the energy density in ordinary matter, but was undoubtedly much more significant in the past. Additional matter in some as yet unidentified dark form appears to be present in greater quantities than the visible matter by a factor of about 10, but it makes itself felt only by its gravitational effects in galaxies and galaxy clusters. The currently most popular hypothesis is that this unseen "dark matter" consists of free weakly interacting elementary particles of a type not yet detected directly on Earth. A final significant contribution to the content of the Universe may come from the energy density of empty space. The possibility that the vacuum itself might influence the gravitational evolution of the Universe was recognied by Einstein, who noted that his equations of General Relativity allowed this possibility through a term involving an additional constant of Nature, the so-called Cosmological Constant.

The long-term evolution of the Universe depends on how the relative contributions of these different components compare with its energy of expansion. If the Cosmological Constant is zero, then both the geometry of the Universe and its long-term future depend on whether total amount of normal and dark matter exceeds a critical value set by the current expansion rate. If it does, then the Universe is positively curved, finite in volume, and will reverse its expansion at some time in the future, collapsing back towards a "Big Crunch". If the amount of matter lies below the critical value, the Universe is open, infinite, and will expand for ever. On the other hand, if the Cosmological Constant is non-zero, as some recent observations seem to suggest, then the situation is more complex. For example, the total content of normal and dark matter may be well below the critical value and yet the Universe may have a flat or even a closed geometry. This is the current "best bet" model among astronomers. It will expand forever, driven by the pressure associated with he Cosmological Constant (and so with empty space) regardless of its geometry.

The third aspect of the Standard Cosmology is a theory for the origin and nature of the structure which is superposed on the homogeneous and isotropic FRW world-model and which ultimately gives rise to galaxies and stars, planets and people. For a long time it was thought that this structure was imposed on the initial state of the Universe at the Big Bang, that galaxies and galaxy clusters were already present in latent form at the creation event. More recently two physical explanations for the origin of structure have been proposed which assign it to processes occurring after the Big Bang during the early instants of the cosmic expansion. The most popular hypothesis is that the Universe went through an early period of accelerated expansion as its matter content underwent a phase transition. During this "inflation", quantum fluctuations in the matter field driving the expanison are blown up to very large spatial scales and so can be identified as the seeds for the later formation of galaxies. Thus the origin of the largest objects known is thus traced back to quantum effects on the smallest scales and the highest energies to which current physics has so far been applied.

The usefulness of this idea lies in the fact that it makes definite predictions for the statistical properties of the seed fluctuations, and so, in principle, for those of observable structures. This turns it from a interesting conjecture into a physically motivated hypothesis open to direct observational test. I discuss below how such tests are carried out in practice. For the time being I note only that quantum effects during inflation appear capable of reproducing many of the observed properties of the present matter and radiation distributions. The second recent hypothesis for the origin of structure supposes that an early phase transition in the matter content of the Universe produced defects in the same way that the crystallization of molten quartz produces dislocations which separate individual pure crystals. Cosmological defects can produce gravitational perturbations on scales comparable to or larger than those of galaxies, and so are also candidate seeds for present structures. Again the statistical properties of the seeds, and so of present structures, can, in principle, be computed. Furthermore, these predictions differ from those of the inflationary model. Currently, they appear difficult to reconcile with observation, and as a result the defect model has fallen out of favour.

The principal ingredients of the standard theoretical model are thus the geometry and the evolution of the FRW models, which derive from General Relativity under the strong assumptions of the Cosmological Principle, a physical content of normal and dark matter, together with radiation and perhaps a Cosmological Constant, and structure which originated from quantum effects during early inflation, or possibly from cosmological defects. At early times the temperatures and densities in such models are very high and the radiation is the dominant component, hence the traditional name, the Hot Big Bang. Let me now move to a description of the observational evidence which leads us to believe that this picture may be more than another Creation Myth, perhaps even a simplified image of "true" cosmic evolution.

3) The Pillars of the Big Bang

The theoretical edifice just described rests on a relatively small number of observational pillars. The first, of course, is Hubble's discovery that the spectra of galaxies show a redshift which increases in proportion to their apparent distance from us. Hubble himself initially interpreted this redshift as a Doppler effect produced by a uniform expansion of the entire universe of galaxies, but towards the end of his life he had doubts and began to consider other possible explanations. More recent work has dramatically reinforced his original interpretation and has ruled out almost all suggested alternatives. Thus Hubble's discovery was shown that on the largest scales our cosmos is not static but dynamic. The 1940's and 50's saw a lively debate about whether the Universe is eternally unchanging but with continuous creation off-setting the dilution due to expansion, or continuously evolving after a single instantaneous creation event. As already noted above, this has now been resolved decisively in favour of the second possibility as a result of other observational evidence which I discuss next.

In the early part of this century it was realised that all astronomical bodies larger than individual stars have almost the same chemical composition. Almost three quarters of their visible mass is hydrogen, just under one quarter is helium, and at most 2% consists of all the other chemical elements put together. In the mid-1940's George Gamow realised that the predominance of hydrogen and helium and the uniformity of their relative abundance could be understood if the Universe had once passed through a hot and dense phase.

Further he was able to calculate that the 3-to-1 ratio would result if the Universe was dominated by radiation at early times, and if the residual radiation today had a thermal spectrum with temperature of about 5 degrees above absolute zero. Such radiation was undetectable when he made this prediction, and the prediction had been forgotten (and then remade) by the time the microwave background radiation was discovered in 1963.

The real cosmic microwave background has a temperature of 2.73 degrees above absolute zero and its spectrum has been shown by the COBE satellite to agree to extraordinarily high precision with Planck's theoretical prediction for a radiation field in complete thermodynamic equilibrium. This is now the strongest piece of evidence arguing that our Universe was once dense, hot and nearly uniform. Only under such conditions could the radiation now arriving from different directions have come so close to equilibrium and at the same temperature everywhere. With microwave telescopes we thus detect directly the afterglow of the Hot Big Bang. Modern updates of Gamow's nucleosynthesis calculations show that both the measured abundances of the light elements and the temperature of the background radiation can be accommodated in the standard model provided the total amount of normal matter is a few percent of the critical value required to close the Universe. This is much less than the amount of unseen dark matter required to induce the observed deviations of galaxy motions from a pure uniform expansion. This is the main direct argument in favour of most of the dark matter consisting of weakly interacting elementary particles of some entirely new type.

COBE's second major contribution in support of the standard cosmology was its demonstration that although the temperature measured for the microwave background radiation is very nearly the same in all directions, significant variations from point to point on the sky are nevertheless detectable at a level of a few parts in 100,000. Such fluctuations are expected in the standard model because the seeds for structure formation were already present at the time when the thermal radiation last interacted significantly with other matter. (This was about 300,000 years after the Big Bang when electrons, protons and helium nuclei first combined to make neutral atoms.) At this time the seed fluctuations imprinted their structure on the background radiation which then propagated unimpeded until COBE observed it. It was a triumph for the standard model that the fluctuation level measured by

COBE was within a factor of 2 or so of that predicted before the mission was launched. We can now claim to have measured some properties of the structure present in the Universe when it was only 300,000 years old. The next challenge is to see if this structure appears compatible with that we observe around us today.

4) Computers as Cosmic Accelerators

COBE's observations were able to provide only a rather crude map of fluctuations in the microwave background. Over the next decade new experiments will enormously improve this situation and will provide a detailed picture of cosmic structure 300,000 years after the Big Bang. Although deviations from uniformity were very weak at that time, we will soon have a very precise characterisation of their statistical properties. The best information about structure in the present Universe comes from large surveys which map the galaxy distribution, bringing to light a large-scale pattern of clusters, filaments and walls. The standard model of cosmology would clearly gain enormous support from a convincing demonstration that the matter distribution inferred from microwave background observations necessarily evolves into one which matches this present-day pattern. Unfortunately, although the standard model supposes the principal physical process driving this evolution to be gravitational instability and, in principle, to be well understood, too many complicating factors influence the formation of the observed structures (not least, the details of how individual galaxies form) for a reliable mathematical treatment of the evolution to be possible. Furthermore, evolutionary timescales are far too long for the changes to be observed directly. In this situation the most successful strategy has proved to be direct numerical simulation of cosmic evolution.

In a curious way numerical simulation plays a role in astrophysics and cosmology which has analogies to that of experiment in other physical sciences. Thus a cosmologist can use some specific model for the the origin of structure, consistent with what we know from studies of the microwave background, to construct a computer representation of the matter distribution 300,000 years after the Big Bang in a "typical" region of the Universe. The computer can then follow how inhomogeneities in this matter distribution are amplified by gravitational forces as the artificial universe grows to the age of the real Universe. The result can be compared to the pattern of structure seen in surveys of the galaxy distribution. The advantage of this approach is that it accelerates many billions of years of cosmic evolution so that they can be modelled in a few days of machine time. The disadvantage is that the computer experiment tells us only about the behaviour of solutions to the particular set of model equations represented in the computer program, whereas normal physics experiments tell us something about the behaviour of the natural world.

Despite this fundamental limitation, numerical simulations have come to play a dominant role in studies of structure formation. This is in part because a simulation predicts all the properties of the matter distribution in some typical region, and so can be compared very directly with real galaxy surveys. Both qualititive comparisons based on pictures and quantitative comparisons based on a variety of statistical measures of structure have proven very useful. At present a substantial majority of cosmologists is convinced that some version of the standard model can be consistent both with the structure seen in the microwave background radiation and with that seen in the distribution of nearby galaxies. Simple treatments of the formation of galaxies in this model have shown that the evolution predicted between these two epochs is qualitatively similar to, and may be quantitatively consistent with, that seen when very distant (and hence young) galaxies are compared with those seen nearby. It is interesting that these structure formation models predicted a number of the "surprising" properties of the distant galaxy populations well before these populations were discovered. Although such models are highly simplified and are based on relatively few independent observational pillars, they have significant predictive power and can be falsified. Results from the new generations of microwave background experiments and of galaxy surveys should be precise enough either to provide a dramatic confirmation of these theoretical ideas or to exclude them.

5) The Inflation Paradigm

While the evolution of structure from the epoch seen in the microwave background to the present day can be followed in detail only by direct numerical simulation, evolution at earlier times is simpler precisely because the deviations from uniformity are then so small. In a model where structure is a result of quantum fluctuations occurring during an early period of inflation, it is possible to give an almost exact mathematical treatment of the evolution of structure from its generation until the microwave background epoch. This is, perhaps, the main reason why cosmologists are excited by the prospect of getting precise measurements of the properties of microwave background fluctuations. If these agree in detail with the predictions of the inflationary model they will not only give very strong support to that model but will also provide detailed information on certain aspects of physics at much higher energies and at much earlier times than have ever been probed before.

The idea of an early period of inflation was not originally proposed as a solution for the origin of structure, but rather as a means for understanding a number of other apparently surprising aspects of our Universe. These puzzles are related to its large size and age, to its large-scale homogeneity and isotropy, and to the fact that its total matter content appears to be relatively close to the critical value which determines whether or not it could be spatially infinite. In FRW world-models without inflation all these properties seem to require the initial conditions just after the Big Bang to be set in an extraordinarily finely balanced and apparently arbitrary way. Thus the evolution becomes deterministic about 10^{-40}seconds after the Big Bang at which time the natural units for measuring its size are of order 10^{-30}cm. It seems fantastic and unnatural that things should "just happen" to have been set up in such a way that the Universe survives for 10 billion years without either recollapsing or going into free expansion and with a curvature radius which is observed to larger than 10^{28}cm. The observed isotropy of the microwave background also appears inexplicable because in these models the regions seen in different directions have never been in causal contact and so have no way to have "arranged" to be at almost precisely the same temperature.

Inflation "solves" the size, age and curvature problems because the radius of curvature of the observed region of our Universe can be increased to very large values during the period of accelerated expansion, resulting in survival of this region until late times. The isotropy problem is also "solved" because the regions which emitted the microwave radiation seen coming from different directions were sufficiently close before inflation to have equilibrated with very similar properties, thus ensuring that they are still very similar at the time they are observed. A natural consequence of this line of argument is that the radius of curvature of the observable region of the Universe should be much larger than its current size. This region would thus appear to be geometrically flat, requiring the total energy content in matter, radiation and cosmological constant to be very close to the critical value for closure.

Unfortunately, although the relatively simple models for inflation, worked out when these ideas were first proposed, do indeed predict that our Universe should appear flat and that structure resulting from quantum fluctuations should have simple and well defined statistical properties, later work has shown that a wide range of more complex physical models can also give rise to inflation and that in some of these the present Universe can appear both isotropic and open, while in others the statistical properties of the fluctuations can differ substantially from those originally predicted. Thus although a flat universe with simple fluctuation statistics appears the most natural prediction of inflation, it is not a requirement. An observational measurement of a very large radius of curvature for our Universe or of microwave background fluctuations consistent with the simple model would provide strong support for the inflationary paradigm, but observations in conflict with either of these would not rule it out. This is not entirely satisfactory – if no conceivable observation could rule out inflation, then it is unclear that the concept has any real physical content. As we shall see in the next section there are alternative lines of argument which can also, perhaps, explain some of the puzzles addressed by inflation.

6) Anthropic Arguments – A Human Role in the Cosmos

Once Hubble had established the cosmic expansion, it was easy to estimate from his measurements approximately how long the Universe takes to double in size. This produced a major surprise – the result is quite close to the estimated age of the Solar System. This apparent coincidence is obviously unexpected in the Steady State Theory, which posits an eternal Universe where the physics determining the expansion rate has no relation to that determining the ages of stars, but it is also surprising in a Big Bang model since, although the Earth and Sun cannot be older than the Universe, there seems to be no reason why they should not be much younger.

A striking suggestion for understanding this is the following. Life as we know it is based on the chemistry of carbon. Carbon was not made in the Big Bang but through processes associated with the evolution of stars. These have a natural timescale of a billion years or more. Thus the raw material for carbon-based life was not available until a billion years after the Big Bang. Furthermore, our current understanding of evolution suggests that similar lengths of time are needed for complex life-forms to develop even under optimal conditions such as a planetary surface partially covered with liquid water. Hence a carbon-based life-form could not measure the expansion rate of the Universe, and so its age, unless its planetary system was formed at least a billion years after the Big Bang and had a central star with a luminous lifetime exceeding a billion years. From this point of view the Universe has to be almost as old as we measure, otherwise we would not be around to measure its age! An additional argument is needed to explain why the Universe is not much older. It appears that the major epoch of star formation is already over – the Universe will make fewer stars like the Sun in the future than it has in the past. Thus if life arises on a "typical" star of solar type, it will most likely find itself in a Universe which is a few billion years old.

This argument, that the Universe must be the way it is because otherwise we would not be around to discuss it, is known as the Anthropic Principle. It arouses strong feelings among scientists because it contrasts dramatically with traditional scientific arguments which ascribe the observed properties of the World to the action of physical laws. Compare, for example, the "explanation" of why the Universe is so old in the last paragraph with the one based on the inflation paradigm which I gave in Section 5. Anthropic arguments also differ fundamentally from the philosophical point of view that our description of the world, hence the world as we perceive it, is a reflection of our own internal make-up rather than of any external reality. Some further examples of anthropic reasoning may make this difference clear.

At the end of the 19th century a major disagreement arose between geophysicists and evolutionary biologists on the one hand, and astrophysicists on the other. Geophysicists estimated that hundreds of millions of years were needed to build and erode mountains, and the evolutionary biologists estimated that similar amounts of time were required to explain the diversity and complexity of species. The astrophysicists, however, argued that the Sun's measured energy output would use up both its gravitational binding energy and its stored chemical energy within a few million years. The Solar System could not, therefore, be older than this. Gradually improving calculations on both sides only succeeded in sharpening the disagreement, and in 1899 the geologist Thomas Chamberlain concluded that the problem must lie in the assertion that only chemical and gravitational energy is available to the Sun: "What the internal constitution of the atoms may be is yet open to question. It is not improbable that they are complex organisations and seats of enormous energies. ... No careful chemist would probably be prepared to affirm or deny that the extraordinary conditions at the center of the sun may not set free a portion of this energy." Thus the conviction that physics must allow a Universe where the Earth and its biosphere can exist led Chamberlain to predict that nuclear reactions power the stars. This was well before the nature of atomic nuclei was understood.

The theory of element formation in stars was worked out in the 1950's. While working on this programme, Fred Hoyle noticed a difficulty in making the first elements heavier than helium. This was supposed to involve three helium nuclei fusing together to make a carbon atom. Unfortunately, a detailed calculation suggested that this process is so slow that that negligible amounts of carbon would be produced. Hoyle realised that the situation could be saved if the carbon nucleus happened to have an excited energy level very near 7.7 MeV; in this case the formation of carbon would occur primarily to this level and would be greatly enhanced. Although the internal structure of the carbon nucleus was unknown when this prediction was made, nuclear physicists were quite capable of testing it. Within a few weeks the predicted level had been found. Reflecting on this later, Hoyle noted that three coincidences are necessary for stars to produce significant carbon: the unstable isotope Be^8, formed from two helium nuclei as an intermediate step, must have an unusually long half-life if it is to capture the third helium nucleus before it decays; carbon must have the right excited nuclear state; and oxygen must have no excited state near 7.2 MeV, otherwise each newly formed carbon nucleus will capture another helium nucleus and turn into oxygen. Only if the constants of nuclear physics have very specific values will all these coincidences hold, and only then will stars be able to manufacture elements heavier than helium. In *our* Universe, therefore, the constants must take these values.

As a final example consider the Cosmological Constant. This additional constant in Einstein's theory of gravity is often considered one of the deepest puzzles of current physics. Its "natural" value can be estimated by requiring gravitational and quantum effects to be comparable in size. Observations of distant galaxies show, however, that its actual value is smaller than implied by this estimate by at least 120 powers of ten. This is the worst known example of what physicists call a fine-tuning problem – a parameter turns out to have a very specific and special value to high precision even though no theory requires this. Such problems are often resolved by the discovery of new physical principles (for example symmetry principles) which force the special value observed. In the present case, for example, such a principle might require the Cosmological Constant to be exactly zero.

Here again an anthropic argument can offer an alternative explanation. Imagine a universe similar to our own in all respects except that the Cosmological Constant takes a value 1000 times greater than the current observational upper limit. Before it is one billion years old such a universe would either stop expanding and recollapse on itself (if the constant is negative) or begin a phase of eternal accelerating expansion (if is is positive). In either case galaxies and normal stars could not form, and as a result life as we know it could not develop. As human beings, we could never observe a universe where the Cosmological Constant is very much larger than the upper limit we currently have on its value.

Adopting such reasoning as part of the standard scientific method is highly controversial and would seem to force some equivalent of a "many worlds" approach to cosmology. If, for example, our Universe is infinite and inhomogeneous on very large scales, then one can imagine that the constants of Nature are not absolute but are set in some dynamic way which causes them to vary slowly over scales much larger than that of our presently observable Universe. Each region will then evolve like an independent universe and will have different apparent values for the physical "constants", at least until it is much older than current estimates of the cosmic age. Only those regions where the Cosmological Constant and the constants of nuclear physics take appropriate apparent values will produce stars capable of synthesizing carbon and heavier elements. All carbon-based life-forms will therefore measure similar apparent values for these constants. In this interpretation human nature determines not only how we describe the observable Universe, but also its physical structure.

Bibliography

Silk, J. 1989, The Big Bang, Freeman, New York.

Barrow, J.D. and Tipler, F.J. 1986, The Cosmological Anthropic Principle, Clarendon Press, Oxford.

Linde, A.D. 1990, Inflation and Quantum Cosmology, Academic Press, Boston.