Part I

Issues in the Philosophy of Cosmology

1

The Domain of Cosmology and the Testing of Cosmological Theories

GEORGE F. R. ELLIS

This chapter is about foundational themes underlying the scientific study of cosmology:

- What issues will a theory of cosmology deal with?
- What kinds of causation will be taken into account as we consider the relation between chance and necessity in cosmology?
- What kinds of data and arguments will we use to test theories, when they stretch beyond the bounds of observational probing?
- Should we weaken the need for testing and move to a post-empirical phase, as some have suggested?

These are philosophical issues at the foundation of the cosmological enterprise. The answer may be obvious or taken for granted by scientists in many cases, and so seem hardly worth mentioning; but that has demonstrably led to some questionable statements about what is reliably known about cosmology, particularly in popular books and public statements. The premise of this chapter is that it is better to carefully think these issues through and make them explicit, rather than having unexamined assumed views about them shaping cosmological theories and their public presentation. Thus, as in other subjects, being philosophical about what is being undertaken will help clarify practice in the area.

The basic enterprise of cosmology is to use tested physical theories to understand major aspects of the universe in which we live, as observed by telescopes of all kinds. The foundational issue arising is the uniqueness of the universe [66, 27, 28]. Standard methods of scientific theory testing rely on comparing similar objects to determine regularities, so they cannot easily be applied in the cosmological context, where there is no other similar object to use in any comparison. We have to extrapolate from aspects of the universe to hypotheses about the seen and unseen universe as a whole. Furthermore, physical explanations of developments in the very early universe depend on extrapolating physical theories beyond the bounds where they can be tested in a laboratory or particle collider. Hence philosophical issues of necessity arise as we push the theory beyond its testable limits. Making them explicit clarifies what is being done and illuminates issues that need careful attention.

The nature of any proposed cosmological theory is characterized by a set of features which each raise philosophical issues:

1. Scope and goals of the theory
2. Nature of the theory: kinds of causality/explanation envisioned
3. Priors of the theory: the range of alternatives set at the outset
4. Outcomes of the theory: what is claimed to be established or explained
5. Data for the theory: what is measured
6. Limits to testing of outcomes
7. Theory, data, and the limits of science

The following sections will look at each in turn.

1.1 Scope and Goals of the Theory

I distinguish here between the physical subject of cosmology, and the wider concept of cosmologia, which also entertains philosophical issues.

1.1.1 Cosmology

I define "cosmology" as theory dealing with physical cosmology and related mathematical and physical issues. Specifically, it proposes and tests mathematical theories for the physical universe on a large scale, and for structure formation in that context. It is a purely scientific exercise, relating to the hot big bang theory of the background model of cosmology, probably preceded by an earlier phase of exponential expansion, to theories of structure formation in this context, and to the many observational tests of these proposals (see e.g. Silk [88], Dodelson [24], Mukhanov [69], Durrer [25], Ellis, Maartens and MacCallum [31], Peter and Uzan [81]). It is a very successful application of general relativity theory to the large-scale structure of spacetime in a suitable geometrical and physical context. However, because of the exceptional nature of cosmology as a science, it pushes science to the limits, particularly when considering what can be said scientifically about the origins of the universe, or existence of a multiverse.

1.1.2 Cosmologia

By contrast, "cosmologia" considers all this, but in addition deals in one way or another with the major themes of the origin of life and the nature of existence that are raised because the physical universe (characterised by "cosmology") is the context for our physical being. It does so in a way compatible with what we know about physical cosmology and physics. Cosmologia necessarily considers major themes in philosophy and metaphysics, perhaps relating them to issues of meaning and purpose in our lives, and is of great public interest and concern.

1.1.3 Which are we Studying?

The major point to make here is that theorists tackling cosmological issues need to make clear which topic they are dealing with: cosmology, or cosmologia. Either is a legitimate topic for investigation, and one can choose which to tackle, but it should be very clear which is the topic of discussion. What is not legitimate is to use only the methods of cosmology, and claim to solve problems of cosmologia. If one wants to tackle cosmologia, one must use adequate methods to do so, involving an adequate scope of enquiry, methods, and data, as discussed further below.

Now one may think the latter ("cosmologia") is not what academic cosmologists should be dealing with; and indeed most groups dedicated to cosmology in physics or astronomy departments restrict themselves to cosmology as defined here. However, some scientists studying cosmology in such departments are proclaiming confidently about issues of cosmologia in highly publicized books and lectures (e.g. Susskind [92], Hawking and Mlodinow [52], Krauss [57]). Thus in these cases the drive to consider these broader issues comes from cosmologists themselves.

The theme of this section can be stated as:

Issue 1: *We can legitimately consider either the enterprise of cosmology as defined here, or extend our investigations to the further issues referred to in cosmologia; but we must make a clear decision as to which stance we are taking, and then adhere to this clearly and consistently in our work. The scope of the theory proposed should be made very clear at the outset.*

In this way, workers in the field can make quite clear if they are engaged in a purely scientific enterprise of cosmology, or are also entering the philosophical and metaphysical waters embodied in cosmologia, commenting on issues of meaning and purpose as well as discussing issues in physical cosmology as evidenced by astronomical observations.

1.2 Kinds of Causality/Explanation Envisioned

Given a statement of scope of the theory, the next issue is what kind of causality will be taken into account. What kinds of causal influences are assumed to be possible in the universe? How do they relate to the great issues of chance, necessity, and purpose?

Assuming we are dealing with cosmology rather than cosmologia, the basic underlying assumption is that what happens at the cosmic scale is the outcome of the interplay of chance and necessity alone. The fundamental problem is that in the cosmological context of a unique universe, the difference between them is blurred.

1.2.1 Necessity

The inexorable nature of necessity is taken to be embodied in fixed and immutable physical laws, expressed in mathematical form [76]. They are taken to be eternal and unchanging, being the same everywhere in the universe at all times and places. The nature of the causal laws proposed (variational principles, symmetry principles, equations of state, and so on) characterizes the causal factors in action in the physical universe that underlie necessity.

Sometimes it is suggested that the laws of physics change with time or place, for example through variation of constants such as the gravitational constant G [96], or choice of a string theory vacuum; but if so, a higher set of laws will be proposed that determine how this happens, for example laws determining how G changes with time, or the laws of string theory that show how string vacua affect local physics. If this is not done, we have no ability to predict physical outcomes scientifically. If this is done, then it is this higher-level set of laws that are the fundamental unchanging ones that govern what happens: in essence one's first stab at finding the unchanging laws was wrong, but that does not mean they do not exist: the higher level set are of this nature.

The Nature of the Laws of Physics

The laws of physics are not physical entities; they are abstract relations characterizing the behavior of matter or fields. It is not clear if the laws are prescriptive, controlling what happens, or descriptive, describing what happens. If they are prescriptive, where or in what way do they exist? How do they get their causal power? If they are descriptive, something else controls how matter behaves: what is it, and how does it get its causal power? And then, why does matter everywhere behave in the same way, so that it is described by mathematical laws?

Possibility Spaces

It is not clear how to obtain traction in considering these issues. A proposal that to some degree sidesteps them is the idea that possibility spaces are the best way to describe the effects of physical laws. Deterministic laws $\mathcal{L}$ are associated with possibility spaces Ω_P that delimit their possible outcomes, and express the nature of necessity by characterizing what is and what is not possible. In effect, we characterize causality not by the nature of the laws themselves, but by the nature of their solution spaces, which strictly constrain what is possible in the physical world. These include phase spaces in classical theory, Hilbert spaces in quantum theory, the landscape of string theory, and so on. Possibility spaces are hierarchically structured, with multiple layers of description and effective behavior depending on the level of averaging used. Constraints such as conservation laws characterize allowable trajectories in these spaces, and so largely define the dynamics (e.g. in Hamiltonian systems).

1.2.2 Ontology and Epistemology

A distinction is crucial here. Possibility spaces themselves exist as unchanging abstract (Platonic) spaces Ω_P limiting all possible structures and motions of physical systems. They are the same at all places and times. Our knowledge of them however is a representation of that space that is changing with time. That is, we represent Ω_P by some projection:

$$\mathcal{P}(t) : \Omega_P \to E_P \tag{1.1}$$

into a representation space E_P where $\mathcal{P}(t)$ depends on the representation we use, and changes with time. This does not mean that physics itself, or the possibilities it allows, are changing: it is just that our knowledge of it is changing with time. Ontology (what possibilities exist, as a matter of fact) is entailed by the nature of Ω_P. Epistemology (what we know about it) is determined by the projection $\mathcal{P}(t)$. The representation space E_P will be represented via some coordinate system and set of units, which can be altered without changing the nature of the entities being represented. Such coordinate changes therefore represent symmetries in the space of possible representations.

The Nature of the Laws

The fundamental issue for cosmology is two-fold. Firstly,

- What kinds of causal laws $\mathcal{L}$ and associated possibility spaces Ω_P hold in the universe? What are their properties?

That is a scientific issue, determined by experiment and observation. According to our current understanding, the laws of physics are locally describable in terms of suitable mathematical equations. They will involve:

- a description of the variables involved and their interactions, governed by specific charges and masses,
- associated variational principles, leading to dynamic equations,
- symmetry principles and associated conservation laws and constraints,
- a specification of the geometry and number of dimensions of the space in which this all happens.

This leads to appropriate partial differential equations, such as Maxwell's equations, Yang–Mills equations, the Einstein field equations, the Dirac equation, and so on, that can be used to calculate the outcome of the application of the laws, given suitable initial data. There will be alternative ways of expressing these laws, for example it may be possible to express them in Hamiltonian or Lagrangian form, or as path integrals, or as partial differential equations, or as integral equations. Given suitable constraints on the physical situation considered, the partial differential equations may reduce to families of ordinary differential equations for suitable variables, which can be expressed in dynamical systems terms.

Why do they have their Specific Nature?

Secondly,

- What underlies the existence of the specific laws $\mathcal{L}$ that hold and associated physical possibility spaces Ω_P? Why do they have the nature they have?

This is a philosophical issue, because there is no experiment we can do to test this feature: we can test what characteristics they have, indeed this is what we do when we determine the effective laws of physics, but we cannot test why they have this nature or what underpins their existence. As far as physical cosmology is concerned, the standard assumption is that their nature, having been determined in some unknown way before the universe began, cannot be different. They just are what they are.

1.2.3 Contingency

Given the laws of physics, the outcome depends on a set of boundary conditions $\mathcal{B}$ for the relevant equations. These are contingent in that, unlike the laws of physics, they could have been otherwise. This is the essential difference between laws and boundary conditions.

Then one has to ask what fixes the specific boundary conditions that actually occurred. In the past, the assumption was that it was fixed by a symmetry principle (the "Cosmological Principle") [11]. The current tendency is to assume they are fixed by chance, that is, some random process determines them [46]. The outcome (the universe that actually comes into being) is then fixed by the deterministic laws $\mathcal{L}$ that map Ω_P into itself, giving a unique outcome from the initial data:

$$\mathcal{L}(\mathcal{B}) : \Omega_P \to \Omega_P \tag{1.2}$$

with different outcomes for different initial data $\mathcal{B}$.

1.2.4 The Relation Between Them

But what does contingency mean in the context of the universe as a whole? Is it a meaningful concept? If so how do we cash it out? The difference is clear in the case of systems situated in the universe, but problematic for the universe itself.

The problem is that as we have only one universe to observe, so we have only one set of initial conditions to relate to; and no way to test if they could indeed have been different. All we experience is that one specific set of initial conditions has indeed occurred. It could be that only one set of initial conditions is possible: in which case their value is fixed by a law, not by a choice among a number of possible different initial states, which is the essence of the idea of contingency.

Are there laws for the cosmos as a whole? This is highly disputed territory. In the 1960s–1970s the idea of a Cosmological Principle was proposed [11], which in effect is a law of initial conditions for cosmology, determining what spacetimes actually occur out of all those that are possible. It determines the subspace $\mathcal{C}$ of realizable cosmologies out of the set of all possibilities Ω_P. Note that this is still an abstract set of possibilities, as it will be a family of space-times rather than only a specific one with all parameters determined, assumed to lead somehow to one or other actually existing universe out of these possibilities. More recently the proposal that there is a law of initial conditions for the universe as a whole has been made by some authors on the basis of various ideas about quantum cosmology (see Hartle [48]). But the concept of a law that applies to only one object is in conflict with the essential nature of a law, namely that it applies to a class of similar objects. To give it meaning one in effect has to introduce an actually existing ensemble $\mathcal{E}$ of universes out of those in $\mathcal{C}$, thus denying the uniqueness of the universe, together with an explicit or implicit probability measure $\mathcal{M}_E$ on this ensemble, allowing one to talk about chance in this context.

Three problems occur. First, these proposals are plagued by infinities that tend to occur in such ensembles. Thus the outcome of using a proposed measure may be ill defined. Second, while we can argue for specific such measures $\mathcal{M}_E$ on various grounds, we cannot test any proposed measure by observation or experiment, because we cannot check the outcomes as regards numbers of universes that occur with the predicted frequencies. It is therefore an untestable choice, with the outcome determined by *a priori* philosophical assumptions (such as a "principle of mediocrity" [99]). In practice, measures on the spaces $\mathcal{E}$ or $\mathcal{C}$ of cosmological models are chosen to give desired results: specifically, trying to make the one universe we can see appear to be probable. Finally, one has to clarify if this measure is proposed as relating to a physically existing ensemble of universes, or a conceptual one. If the former, how do we show it exists? If the latter, how does it influence what actually happens?

Issue 2: *How can there be a meaningful difference between chance (i.e. contingent boundary conditions for cosmology) and necessity (the deterministic physical laws that act in an inevitable way, as characterized by possibility spaces) in the context of the existence of the unique single universe we probe by astronomical observations and physical experiments*?

The distinction seems rather arbitrary. We cannot establish the difference observationally or experimentally. Nevertheless we usually assume there is such a difference, as that is how ordinary physics works.

1.2.5 Creation of the Universe

These issues come to a head in terms of theories of creation of the universe (e.g. Hartle and Hawking [49]). The problem with theories of creation of the universe is that we can only proceed by applying physical ideas that we determine within the universe, and extrapolate them to applying to the universe itself. It is not clear that makes sense, *inter alia* because there is no concept of space or time before the universe exists, and certainly there is no way to test their validity (we cannot rerun the project and try with varied starting conditions, for example).

To additionally propose that one has a theory of creation of the universe "from nothing" (e.g. Krauss [57]) does not make scientific sense, for one can only develop a creation theory by assuming the nature of the laws to be applied: quantum field theory, for example [80], perhaps extended to the standard model of particle physics, or something like it. These are assumed to causally precede the coming into being of the physical universe. Calling that "nothing" is sleight of hand [4]: it is a very complex structure that is assumed to in some sense pre-exist the coming into being of the universe (for they are assumed to cause its existence). So where or in what sense do all these items exist prior to the existence of the universe itself? Presumably they are meant to inhabit some kind of Platonic space that pre-exists the universe, which somehow gains causal power over material entities even though space and time do not exist then. That is a very powerful form of existence.

There is a lot that needs clarification here. Supposing this was indeed clarified into a consistent theory, it would not be a testable theory of how things happened, even if some proposed outcomes might be testable.

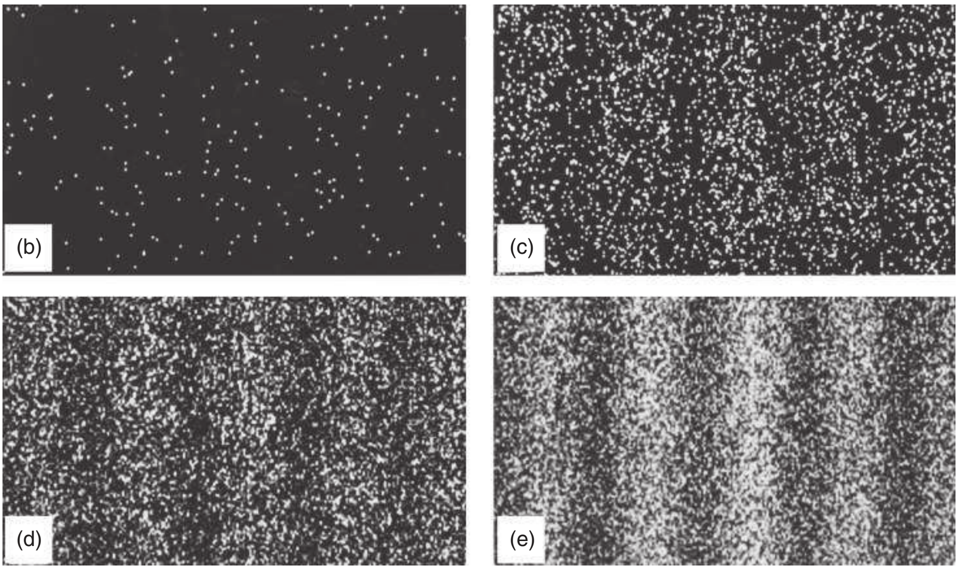

Figure 1.1 **Irreducible randomness occurs in experimental quantum physics**. Double-slit-experiment performed by Dr. Tonomura showing the build-up of an interference pattern of single electrons. The numbers of electrons are (b) 200, (c) 6,000, (d) 40,000, (e) 140,000 [from Wikimedia commons]. Quantum theory predicts the statistics of outcomes that eventually builds up, but cannot even in principle predict where each individual electron will arrive.

1.2.6 Chance: Randomness

Chance occurs within the universe, due to statistical interaction between emergent levels of structure. However, attractors in possibility spaces determine basins of attraction, and so relate chance and necessity: substantial variation of data can lead to the same outcomes, reducing the effect of randomness. In the case of strange attractors, the effect is the opposite: outcomes are unpredictable in practice, even though they are predictable in principle.

Irreducible Quantum Uncertainty

However, in addition, irreducible quantum uncertainty occurs in the universe due to quantum physics effects: the initial data do not determine the outcome uniquely, even in principle, because quantum physics underlies all physics. This is illustrated in the double-slit experiment outcomes shown in Figure 1.1.

Assuming that standard quantum theory as tested by many experiments is right, there is no way even in principle of determining where each individual photon will end up on the screen after passing through a very narrow double slit, although the statistics of the outcomes are fully determined by the Schrödinger equation [37, 55, 108].

Now it is rather strange that many working physicists talk about quantum physics as if it were a deterministic theory. This is because the Schrödinger equation is a deterministic equation, uniquely determining the evolution of the wave function Ψ from initial data. But the value of Ψ does not determine specific physical outcomes: it determines only the statistics of these outcomes [37, 38, 55]. When a measurement takes place, the unitary Schrödinger equation does not apply: a superposition of eigenstates collapses to a single eigenstate (see e.g. Zettli [108]: 158), which is a non-unitary change. The philosophy seems to be that as the behavior of an ensemble of events, and in particular the statistics of associated energy changes and scattering angles, is uniquely determined, quantum mechanics is a deterministic theory. But there is no ensemble to consider without individual events! (as is seen in Figure 1.1): and irreducible uncertainty applies to the individual events. Physics is in fact unable to predict unique specific outcomes from initial data: it does not (at the micro level) determine the particular individual things that happen.

Various attempts have been made to show this is not the case, e.g. the Many Worlds (Everett) interpretation of quantum physics [55], and the Bohm Pilot Wave Theory [8]. In these cases one has a deterministic (unitary) outcome taking place behind the scenes of what is actually experienced by physicists working in a laboratory. None of these proposals alters the fact that experiments definitively demonstrate that we actually experience irreducible quantum uncertainty in a laboratory as we try to predict the outcomes for specific individual events from measurable initial data [37, 108].

Unpredictability in Cosmology

Now one might think that this is irrelevant to physics at a cosmological scale, because this only happens at microscales. However, if the current standard model of cosmology (see Section 1.4) is correct, this is not the case, because in this theory, quantum fluctuations during the inflationary era are amplified to macroscopic scales by the exponential expansion that occurs, and then result in the seed perturbations on the last scattering surface that lead to the formation of clusters of galaxies and galaxies by a bottom-up process (small-scale structures form first, and then assemble to form larger scale structures later).

This means that if we knew everything measurable about the universe at the start of inflation, this would not determine the specific individual astronomical structures that we can observe in the universe at later times. The statistics of these structures is predictable and indeed is used as a test of cosmological theory [24, 69, 25, 31, 81]. The existence of specific individual structures such as our own galaxy is not caused by the quantum processes in operation, for this is the nature of quantum physics as has been conclusively determined by laboratory experiment, as shown in Figure 1.1: it is subject to irreducible uncertainty. Given complete data at the start of the universe, cosmological theory cannot predict the specific structures that will later come into being. The conclusion is that necessity does not always apply in the universe, even at cosmological scales. However, the statistics of what happens is uniquely determined.