# 1 The science of the one universe in time

## THE SINGULAR EXISTENCE OF THE UNIVERSE

This book develops three connected ideas about the nature of the universe and of our relation to it. The first idea is that there is only one universe at a time. The second idea is that time is real and inclusive. Nothing, including the laws of nature, stands outside time. The third idea is that mathematics has this one real, time-drenched world as its subject matter, from a vantage point abstracting from both time and phenomenal particularity.

On the view defined by these three ideas, the universe is all that exists. That there is only one universe at a time justifies using the terms universe and world interchangeably. If there were a plurality of universes, the world would be that plurality. The singular universe must, however, be distinguished from the observable universe, for our universe may be much larger than the part of it that we can observe. In this book, we use the words cosmology and cosmological to designate what pertains to the universe as a whole, not just to its observable portion. Observational astronomy has continued, in recent decades, to make remarkable discoveries about the observable universe. Cosmology, however, risks losing its way. The arguments of this work are cosmological: they concern the whole of the universe and the way to think about it.

Each of the three central ideas developed in this book has implications for how we interpret what science, especially in physics and cosmology, has already discovered about the world and for how we view what science can and should do next. It has consequences as well for our view of the place of these scientific discoveries in our self-understanding.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

The first idea is the solitary existence of the universe. We have reason to believe in the existence of only one universe at a time, the universe in which we find ourselves. Nothing science has discovered up to now justifies the belief that our universe is only one of many, although the universe may well have predecessors. The multiplication of universes in contemporary cosmology has not resulted from any empirical discovery or inference from observation; it has been the outcome of an attempt to convert, through this fabrication, an explanatory failure into an explanatory success. The explanatory failure is the compatibility of a prevalent view of how nature works at the level of its elementary constituents with many states of nature other than the one that we observe. (Today, in the early twenty-first century, string theory, with its prodigious surfeit of alternative consistent versions, almost all of them not realized in the observed universe, provides the most striking example of such underdetermination of phenomena by prevailing theories.) The conversion of failure into success proceeds by the simple expedient of supposing that for each version or interpretation of the theory in favor there is a corresponding universe in which what it says is true.

If these unobserved universes were held to be merely possible, the question would arise why only one of the possible universes in fact exists. Therefore, the most radical form of the conversion of failure into success consists in claiming that these other universes are more than merely possible; they are actual, even though we have no evidence of their existence (the multiverse idea).

The most widely accepted causal hypothesis today to explain the genesis of such a multiverse is "eternal inflation," postulating the creation of an infinite number of universes formed as bubbles from phase transitions on an eternally inflating medium. Within string theory, it is plausible to believe that such bubble universes are described by laws, chosen by a stochastic process from the immense range of theories that are compatible with the string-theoretical approach. The retrospective teleology of the "strong anthropic principle," according to which the criterion of selection of the laws in our

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

universe is that they make possible our human life and consciousness, closes the circle of prestidigitation.

The sleight of hand represented by this combination of ideas amounts to an ominous turn in the history of science. It is a turn away from some of the methods, standards, and presuppositions that have guided and disciplined science until relatively recently.

Although the opposing idea, of the singular existence of the universe, may appear self-evident to some scientists and to many non-scientists, it raises a problem of the first order. Individual being, wrote Aristotle, is ineffable. We can provide law-like explanations of recurrent phenomena in parts of the universe. But how can there be a law-like explanation of the universe as a whole if the universe is one of a kind? How can we offer such an account if we are not entitled to represent and to explain our world as one of many possible or even actual worlds? The theory of the universe would have to be the theory of an individual entity. For such a theory the history of science offers no model.

## THE INCLUSIVE REALITY OF TIME

The second idea defended and developed in this book is that time is inclusively real. According to this thesis, nothing in this singular universe of ours remains outside time.

The reality of time may seem an empty truism. In fact, it is a revolutionary proposition. It contradicts not only certain speculative doctrines that openly affirm the illusory character of time, but also ideas about causation and scientific explanation that may seem beyond reproach and doubt.

When the idea of the reality of time is combined with the idea of the unique existence of the observed universe, it results in the view that this one world of ours and every piece of it have a history. Everything changes sooner or later.

Recognition of the reality of time gives rise to a philosophical conundrum about causation. If time were not real, there could be no causal relations for the reason that there would be no before (the cause)

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

and after (the effect). Causes and consequences would be simultaneous. They would therefore be unreal or mean something different from what we take them to mean. Nothing would distinguish causal connections, which are time-bound, from logical or mathematical relations of implication, which stand outside time. What we, in causal language, call causes and effects would in fact be aspects of a relational grid in a timeless reality.

If, however, everything is time-bound, that principle must apply as well to the laws, symmetries, and constants of nature. There are then no timeless regularities capable of underwriting our causal judgments. Change changes. It is not just the phenomena that change; so do the regularities: the laws, symmetries, and supposed constants of nature.

Our conventional picture of causation must be confused. For we seem to believe, on the evidence of the way in which we use our causal language, outside science as well as within it, that time is real, but not too real. It must be somewhat real; otherwise there would be no causal connections at all. It must not, however, be so real that our causal judgments are all adrift on a sea of changing laws.

In this book we argue that the evidence of science – the deliverances of the science of today, viewed in the light of its recent history – does not entitle us to circumscribe the reality or the reach of time. Our causal judgments cannot indeed be anchored in immutable laws and symmetries. That need not mean, however, that we stand condemned to explanatory impotence. Causal explanation, properly reinterpreted and redirected, can survive the overcoming of our equivocations about the reality of time. It can make peace with the view that time is real and that nothing remains beyond its reach.

This intellectual program brings us face to face with a further riddle, a puzzle that comes into sight when we begin to take seriously the notion that the laws of nature, as well as its other regularities – symmetries and supposed constants – are within time, and therefore susceptible to change, rather than outside time, and therefore changeless. We seem faced with an unacceptable choice between two troubling positions.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

One position is that higher-order or meta-laws govern the change of the laws and other regularities of nature. In this event, however, the problem presented by the time dependence of the laws is simply pushed to the next level. Either such higher-order laws are themselves within time and liable to change, or they are timeless and changeless. Nothing fundamental would have shifted in the structure of the problem.

The other possibility is that no such higher-order laws exist. Then our causal judgments would remain bereft of any apparent basis. The change of laws would seem an enigma for which no adequate explanation can exist: change requires causal explanation, and causal explanation must in turn be warranted, or so it is traditionally believed, by laws and symmetries of nature.

We consider ways out of this dilemma. One of them plays an especially large role in our argument, as it has in the development of the life and earth sciences and of social and historical study, although not of physics. According to this view, the laws, symmetries, and supposed constants change together with the phenomena. Causal connections are, on this view, a primitive feature of nature. In our cooled-down universe, they recur over a discriminate structure of natural phenomena, which is to say that they exhibit law-like form. In other, extreme states of nature, however, those that occurred in the very early history of the universe, they may be, or have been, lawless.

The idea that the laws of nature are susceptible to change and that the laws may develop coevally with the phenomena that we take them to govern may be puzzling: for the reasons that I have suggested, it renders unstable the laws of nature that we habitually take as warrants of causal explanation. However, it is neither nonsensical nor unprecedented. We are accustomed to invoke it in the life sciences as well as in social and historical study. It saves us from needing to appeal to speculative metaphysical conjectures, such as the notion of a multitude of unobservable worlds.

The conjecture of the mutability of the laws of nature seems to give rise to insuperable paradoxes. The impression of paradox, however,

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

begins to dissolve once we turn on its head the conventional picture of the relation between laws and causal connections, and recognize that the former may derive from the latter rather than the other way around. This idea may lead us to think in a new light of a broad range of familiar and intractable facts. Among these facts are the unexplained values of the universal constants of nature, especially of those constants that we do not and cannot use as conventional units of measurement and that are, for this reason, conventionally called dimensionless. Their seemingly arbitrary values may be the result of earlier states of the universe and of the operation of laws or symmetries different from those that now hold. They may be vestigial forms of a suppressed and forgotten history: testimonials to a vanished world – the one real world earlier on.

* * *

A simple way to grasp what is at stake for science in the idea of the inclusive reality of time and of its corollary, the conjecture of the mutability of the laws, symmetries, and supposed constants of nature, is to ask the question: Where do these regularities come from? Because the laws and symmetries of nature, as we now understand them, fail to account uniquely for the initial conditions of the universe, we need to ask as well a second question: Where do these initial conditions come from? (The mysterious constants of nature help describe these conditions. They do not explain them. On the contrary, they require explanation, which the established laws and symmetries fail to provide. Thus, even though we can count the constants, together with the laws and symmetries, as regularities, we cannot expect them to help explain the initial conditions of the universe. They form, from the outset, part of the problem rather than part of the solution.)

There are, broadly, three ways to approach these questions.

A first approach is to say that the laws and symmetries comprise an immutable framework of natural events. They are what they are. If they fail to apply to the very earliest moments of the universe or to explain its initial conditions that must be only because our knowledge of the laws and symmetries remains incipient and incomplete. It is this first approach that, at least until recently, has been ascendant in the

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

history of physics, from Newton to Einstein. It represents part of the intellectual backdrop to the major discoveries of twentieth-century cosmology.

An objection to this approach is that what we already know about the very early universe suggests that the laws and symmetries as we now formulate them could not have held in the extreme conditions that existed then and that may exist again later in the history of the universe: for example, in the interior of black holes. How can we speak of laws and symmetries if, in such extreme states, there is no discriminate structure – no stable repertory of different kinds of things, such as those described by the standard model of particle physics, interacting in ways that laws and symmetries can capture?

Another objection is that the initial conditions of the universe, and therefore its subsequent evolution, seem extremely unlikely if nature is indeed constituted as our laws and symmetries say that it is. Under this first approach, the initial conditions of the universe remain unexplained by the laws and symmetries. The laws and symmetries seem applicable only to a universe that has already organized itself in the ways that are characteristic of the cooled-down universe.

It is true that so long as it resists the temptation to succumb to a rationalist metaphysics science can never show that the universe had to be what it has become. Science must in the end recognize what I here call the facticity of the universe: that it just happens to be what it is rather than something else. The problem with the first approach, however, is that it may prematurely and unnecessarily narrow the field open to causal inquiry. It may mistake nature for a subset of natural processes. It may codify as laws and symmetries how nature works in these familiar variations: those that prevail in the relatively cold and differentiated universe in which we find ourselves.

It is one thing to respect the inability of science to show that the universe must be what it is. It is another thing to reduce science to a body of precise laws, symmetries, and constants that are unable to account either for themselves or for the initial conditions of the universe.

A second approach to the question of where the laws and initial conditions of our universe come from is to take this universe as only one of a multitude. The chief object of explanatory ambition shifts, under this approach, from the laws, symmetries, and constants that happen to prevail in this universe of ours to the laws and mathematical conceptions governing the multitude of universes, of which ours would be only one. It will soon appear that there is hardly any difference in this view between laws and mathematical notions.

The effective laws that have up until now been the chief object of science become, under this approach, simply a variant among many sets of higher-order laws applying to the crowd of universes. The strangeness of the initial conditions of our universe can be discounted as a trait of a universe that happens to be an outlier in the crowd. The detachment of the higher-order laws from the realities of the universe that we observe, and their multifarious content, lend them all the more to marriage with mathematics.

The resulting ideas are not so much physical theories expressed in the language of mathematics as they are mathematical conceptions presented as physical theories. Under such a view, the distinction between laws and initial conditions disappears.

The extreme limit of this idea is the notion that natural realities are nothing but mathematical structures. Because such structures are timeless, so must the states of nature that they comprise be timeless. To each mathematical structure there corresponds a universe, instantiating that structure in all its particulars. Observational surprises reveal mathematical ignorance.

This second approach (whether or not in its extreme form) is an invention of the late twentieth and the early twenty-first centuries. It has been almost entirely foreign to the history of physics and cosmology until the last few decades. It arose by the circumstantial convergence of developments in particle physics, culminating in string theory, with the conjecture of a multiverse and the appeal to anthropic reasoning. It found inspiration and reinforcement in mathematics, given the central role that it assigned to mathematical ideas.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

An initial objection to this approach is at once methodological and moral. It invents imaginary entities – all the other unobserved and unobservable universes (in cosmology) or states of affairs (in particle physics) – to save itself from having to confront, in either particle physics or cosmology, the failure of its theoretical conceptions to account for nature as we encounter it. In this way, it wastes the treasure of science, its enigmas.

A second, related objection is that, by using this stratagem, it inverts the relation of physical science to mathematics and elides the difference between them. If mathematics is a storehouse of ideas about the ways in which pieces of reality may connect with one another, then physics, in this account, becomes the identification of each of these mathematical connections with a physical reality. It is, we argue in this book, a practice resting on a misguided view of the relation of mathematics to science and nature.

A third objection – and the one that will be most telling to a scientist – is that at the end of the day this approach evades the work of explanation. It subsumes the unexplained laws and initial conditions under a vast framework of possible variations of nature, all but a tiny number attributed to unobservable universes and unknown states of affairs.

We develop a third approach. Its working assumption is that the more promising way to explain the regularities as well as the structure of nature, and so too the initial conditions of the one real universe, is to explain them historically. This approach proposes that cosmology complete its transformation into a historical science. It seeks empirical support in the most important findings of the cosmology of the last hundred years: those that have to do with the history of the universe and that have been codified incompletely in the now standard cosmological model. Structure results from history more than history derives from structure.

This third approach has many counterparts in the life and earth sciences as well as in the historical study of human society. However, unlike the other two approaches, it counts on few representatives in the history of modern physics and cosmology.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

It fails to explain away the factitious character of the universe – that the universe just happens to be one way rather than another. However, it vastly enlarges the field of causal inquiry. As a result, it suggests an agenda of empirical research that communicates with the major discoveries that cosmology has made over the last hundred years and continues to make now.

The historical approach, as we here understand and develop it, makes use of each of the three chief claims of this book. It discards the fabrication of imaginary universes in favor of a focus on the one real universe and its history. It takes the reality of time so seriously that it refuses to exempt either the basic structure or the fundamental regularities of nature from susceptibility to change. It wants to put mathematics in its place, as an instrument of physical theory rather than as a substitute for it.

Such an approach raises daunting problems. I have already touched on two of them in this early stage of our argument.

The first problem is that if there is only one universe at a time, we must conceive the seemingly paradoxical endeavor of developing the science of a singular entity. The traditional way of avoiding this problem in cosmology is to scale up: to extend explanations developed to address pieces of the universe into ideas about the whole universe. In the cooled-down universe, with its discriminate structure, exhibiting laws and symmetries, such pieces of the universe – for example, patches of space-time – come in multiple instances conforming to the same regularities. Cosmology relies on the amalgamation of theories about local phenomena.

However, scaling up from piecemeal theories of nature to cosmological conceptions, at least insofar as it relies on a distinction between stipulated initial conditions and unchanging laws, deserves to be resisted. It is just what the argument against the first cosmological fallacy and its Newtonian paradigm – an argument developed later in this chapter – forbids. We must face, without the relief that this procedure offers, the difficulty that the universe, as a reality both unique and historical, presents to science: Aristotle's conundrum about the ineffable character of individual being.

The second problem is that if everything in this one universe, including its regularities and its structure, changes sooner or later, we cannot accept a move that has helped define the path taken by physics and cosmology at least since the formulation of special and general relativity and of quantum mechanics in the early twentieth century. The move is to combine denial of an absolute background of space and time, distinguishable from physical events, with the reaffirmation of belief in a permanent structure of ultimate constituents of nature and in an immutable framework of laws and symmetries. The rejection of these twin ideas requires us to change our view of causality and of the relation of causal connections to the laws and symmetries of nature.

## THE SELECTIVE REALISM OF MATHEMATICS

The third idea central to our argument is a conception of mathematics and of its relation to nature and to science. Mathematics, according to this idea, represents a world eviscerated of time and phenomenal particularity. It is a visionary exploration of a simulacrum of the world, from which both time and phenomenal distinction have been sucked out.

Our causal explanations are steeped in time: the cause precedes the effect. If time were illusory, so would any causal nexus be an illusion. On the other hand, however, if time were real and inclusive to the point of resulting in the mutability of the laws of nature, our causal judgments would lack a stable warrant. Our conventional ideas about causation are confused; they assume that time is real, but not too real.

The relations between mathematical and logical propositions are, however, timeless: the conclusion of a syllogism is simultaneous with its premise. They are timeless, even though we reason them through in time, and use them in the analysis of events in time.

In the philosophical and scientific tradition within which the ideas of the singular existence of the universe and of the inclusive reality of time have remained decisive, mathematics has gained a power that none of the well-known positions in the philosophy of

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

mathematics seem adequately to explain or to justify. The laws of nature appear to be written in the language of mathematics. But why and with what significance? The "unreasonable effectiveness of mathematics" remains a riddle without a convincing solution.

In the history of philosophical ideas about mathematics, two sets of conceptions have come close to exhausting approaches to the solution of this riddle. According to the first set, mathematics is discovery of an independent realm of mathematical entities and relations. According to the second set, mathematics is invention: made-up conceptual entities, manipulated according to made-up rules of inference. The problem is that neither of these approaches to mathematics seems to help explain the applicability of mathematics to the world.

We propose a different view, one that begins from the acknowledgment of the contrast between the temporal character of every causal nexus and the timeless quality of mathematical and logical relations. Mathematics is about the world, viewed under the aspect of structured wholes and bundles of relations, disembodied from the time-bound particulars that make up the actual world: effacement of particularity goes together with denial of time.

The world studied by mathematics is not quite our world, the one real world, soaked through and through in time. Neither, however, is it another world, of eternal mathematical objects, separated from ours by an unbridgeable gulf. It is a proxy for our world, a counterfeit version of it, a simulacrum, distinguished from it because in it everything is denuded of placement in time and of phenomenal particularity.

It is as if our mathematical and logical reasoning represented a Trojan Horse, placed in the mind against the recognition of the ultimate reality of time and difference. However, its selectivity – its disregard for time and particularity – is the source of its usefulness.

Instead of regarding our faculty of mathematical and logical reasoning as a way of overcoming the limits of our natural constitution, we should understand it as a part of that constitution. By enabling us to expand and recombine our ideas of how pieces of the world can connect with other pieces, independently of the particulars of any given time-bound circumstance, this faculty vastly enhances the scope of our problem-solving capabilities. It conferred an evolutionary advantage when we were simpler than we now are, and continues to confer one now that we have become more complicated.

The foundation of this advantage lies in its simplifying approach to the one real natural world, rather than in a direct access to another world of timeless and therefore unnatural objects or to an array of possible worlds that never wore the garment of reality. It is a natural faculty that has nature as its subject. However, it increases its power by virtue of its distinct approach to the particulars of this one time-bound universe of ours.

In its early stages, the relation of mathematics to the world of temporal change and of phenomenal particularity is direct: less by induction than by what Pierce called abduction – an imaginative jumping off from an open-ended series of particulars. Soon, however, the predominant relation of mathematics to nature becomes indirect. We begin to expand the range of mathematical ideas by analogy, without license or even provocation from natural experience. We go, for example, from the three-dimensional space of Euclidean geometry, with its simplification of our sensual experience, to geometries that have no counterpart in our perception. We move from the natural integers by which we count things in the world to numbers useless in counting anything we will ever directly encounter and experience with our senses.

The mathematics that we develop on the basis of this indirect relation to nature, driven by an agenda internal to mathematics itself, may or may not apply to the elucidation of natural phenomena. It may or may not be useful in the work of natural science. There is no assurance that it will be serviceable, although it often is. The ultimate source of its power is that it combines connection to nature with distance from nature.

This power perennially tempts us to succumb to two connected illusions. The first illusion is that we have in mathematics a shortcut to indubitable and eternal truth, somehow superior to the rest of our fallible knowledge. The second illusion is that, as the relations among mathematical propositions are timeless, the world itself must somehow participate in the timelessness of mathematics.

Here we offer an account of mathematics that has no truck with either of these illusions. It is a realistic and deflationary view. It claims that we cannot make adequate sense of the effectiveness of mathematics in natural science by treating mathematics either as the exploration of a separate world of timeless mathematical objects or as the free invention of ideas about number and space that turn out, mysteriously, to be applicable to nature.

We enjoy our mathematical powers for natural reasons. We develop them at first inspired by nature, eviscerated of time and particularity, and then at a distance from the original sources of our inspiration. Mathematics, however, is smaller, not greater, than nature. It achieves its force through a simplification that we can easily persuade ourselves to mistake for a revelation and a liberation.

The view of mathematics as the imagination of a counterfeit version of the world, robbed of time and phenomenal particularity, acquires its full force and meaning only when combined with the other two ideas central to our argument: that there is only one real world and that everything in this world changes sooner or later. One world. Real time. Mathematics is about the one world in real time, not about something else. Instead of trying to find what else mathematics could be about other than the world (there is nothing else), we should be concerned to understand in just what sense it can be about a world to the manifest qualities of which it is so strikingly and willfully blind.

## THE FIRST COSMOLOGICAL FALLACY

There is one real universe. Time is real, and nothing lies beyond its reach. Mathematics has the one real, time-soaked world as its subject matter and inspiration. It is useful to the understanding of this world precisely because it explores the most general features of relations among pieces of the world abstracted from both time and phenomenal particularity.

These three propositions form the axis of the argument of this book. To recognize and to develop the truth that they express, we must reject two fallacies. Each of these fallacies enjoys widespread influence within and outside physics and cosmology. They are closely connected.

Taken together, they summarize much of what is misguided in our received understanding of the discoveries of science.

Call them the two cosmological fallacies. Both of them mistake a part for the whole. They make different but connected mistakes. The first fallacy applies to the whole of the universe methods and ideas that can be successful only when applied to part of it. It is a fallacy of false universality: it treats the whole universe as if the whole were one more part. The second fallacy embraces a view of nature and of its laws that is inspired by the forms that nature takes during part of the history of the universe. It is a fallacy of universal anachronism: it applies to the whole history of the universe ideas that are pertinent only to part of that history. Its view of the workings of the natural world is too parochial to do justice to the metamorphoses of nature.

* * *

The first cosmological fallacy – a fallacy of false universality – applies to the whole of the universe, and therefore to the central problems of cosmology, what we here call the Newtonian paradigm. The Newtonian paradigm is the chief method of explanation that physics and cosmology have deployed since the time of Galileo and Newton. Relativity and quantum mechanisms have not disturbed its ascendancy however much they may have modified its application.

Under the Newtonian paradigm, we construct a configuration space within which the movements and changes of a certain range of phenomena can be explained by unchanging laws. The range of experience defined by the configuration space and explained by the laws can in principle be reproduced, either by being found in another part of the universe or by being deliberately copied by the scientist. The recurrence of the same movements and the same changes under the same conditions, or the same provocations, confirms the validity of the laws.

The configuration space within which changeless laws apply to changing phenomena is marked out by initial conditions. These conditions are the factual stipulations defining the background to the phenomena explained by the laws. The stipulations mark out

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

the configuration space: the space within which laws apply to the explained phenomena. By definition, they are not themselves explained by the laws that explain movements and changes within the configuration space. They are assumed rather than explained.

However, that they perform in a particular part of science the role of unexplained stipulations rather than of explained phenomena does not mean that they cannot reappear in another chapter of scientific inquiry as subjects for explanation. In the practice of the Newtonian paradigm what is stipulation for some purpose becomes the subject matter to be explained for another. That the roles of what is to be explained and what does the explaining can in this way be reversed ensures that we can hope to explain all of the universe, part by part.

The observer stands, both in principle and in fact, outside the configuration space. Conceptually, his relation to it resembles the relation of God to the world, in the Semitic monotheisms – Judaism, Christianity, and Islam: not as creator but as observer. He looks upon it, to use an astronomical metaphor, from the vantage point of the stars. The laws go together with this ideal observer. They govern what happens inside the configuration space. They have, however, no history of their own within that space – or anywhere else.

The laws determine changes or movements within the configuration space. Thus, they can be used to explain events in time. To explain changes of the phenomena, it is first necessary to represent them. The most familiar way in which to do so is to plot them as movements along an axis. Time is converted into space.

The laws are timeless. They have no history. They underlie and justify our causal explanations. They are, however, themselves without explanation. To ask why they are what they are is to pose a question that lies in principle beyond the limits of a natural science conforming to the Newtonian paradigm.

Those whose ideas about the practice of science have been formed in this mold may hope to find in mathematics the beginnings of insight into why the laws are what they are. This conjecture remains, however, no more than a metaphysical speculation with limited practical significance for the conduct of science under the guidance of the Newtonian paradigm.

The first cosmological fallacy consists in the application of this way of doing science to the universe as a whole, which is to say to the problems that are distinctive to cosmology. When the topic is the whole of the universe and its history, rather than a part of the universe, the distinction between law-governed phenomena within a configuration space and the stipulated factual conditions defining that space ceases to make sense. There is no place outside the configuration space for anything else to be; that space has become the entire universe. It is no longer thinkable, even in principle, to prepare or even to discover copies for what we are to explain, now the entire universe, so that we can test the constant validity of the laws.

Deaf to Newton's warning not to feign hypotheses, we may appeal to the idea of multiple, parallel universes in an effort to rescue the cosmological uses of the Newtonian paradigm. If, however, these other universes are, as they must be, causally unconnected with our own, and no light-borne information can travel from them to us, this conjecture will amount to no more than a vain metaphysical fantasy, disguised as science.

The process by which what is the factitious stipulation of an initial condition in one local explanation becomes an explained phenomenon in another is now interrupted. In an account regarding the whole universe and its history, no occasion arises for such a reversal of roles. Thus no hope can be well founded that by accumulating local explanations we slowly approach an explanation of the whole.

The observer can no longer stand outside the configuration space, and claim to adopt the godlike view from the stars; all the stars, and everything around them, are dragged down into the field of explanation. If the laws of nature are somehow exempt from the violent changes that nature undergoes, they must exist on some other plane of reality, in the company of mathematics, as it is understood by mathematical Platonists.

Thus, every feature of the Newtonian paradigm fails when its subject matter ceases to be a region of the universe and becomes the entire universe. The denial of this failure is the substance of the first cosmological fallacy. It results in a series of equivocations that corrupt the practice of scientific inquiry and prevent cosmology from remaining faithful to its vocation to be a master science rather than a sideshow.

A major stratagem by which to dismiss or diminish the implications of seeing the first cosmological fallacy for what it is consists in treating the problems of cosmology as peripheral to the agenda of physics in particular and of natural science in general. The parts of the universe, however, are parts of *the* universe. Our view of the universe and its history has implications for our understanding of all of its parts. If, for example, there is a succession rather than a plurality of universes and if the causal connection between successive universes, although stressed, is never broken, many features of our world may have their origin and explanation in the traits of the very early universe or of universes that preceded them.

An influential variation on the strategy of marginalizing cosmology the better to suppress the embarrassments it creates for established scientific ideas and practices is to represent the earliest moments of the universe as characterized by infinite degrees of temperature and energy. That is precisely what marks a singularity in the strict and conventional sense. Once we cross the threshold of the infinite in the representation of nature (rather than just in the exercise of the mathematical imagination), we can no longer make use of any of the explanatory practices, including the Newtonian paradigm, that we are accustomed to apply to the world of nature that we know. Thus, under the view that the present universe began in a singularity the parameters of which are infinite, rather than in a violent event of extreme but finite parameters, we can attribute to the enigmas of the infinite what are in fact confusions and contradictions resulting from the illegitimate universalization of local explanations. It is as if the jump from the finite to the infinite provided a generic license for ideas that, in the absence of such license, would readily be dismissed as untenable.

## THE SECOND COSMOLOGICAL FALLACY

The second cosmological fallacy – a fallacy of universal anachronism – sees the entire history of the universe from the standpoint of ideas that may be pertinent to only part of that history. It incorporates into our practices of scientific explanation a view of the workings of nature that accounts for those workings only in certain states of nature but not in others. The substance of the second cosmological fallacy is to treat the form that nature takes in the differentiated, cooled-down universe as its sole and permanent form. This model of the workings of nature can then be read backward as well as forward, to earlier and later moments of the history of the present universe, as its one and only mode; hence the mark of universal anachronism.

It is a cosmological fallacy because it can arise only within cosmology and as a result of its most significant discovery: the discovery that the universe has a history. The second cosmological fallacy is thus no mere methodological misstep. It amounts to a misreading of the facts of the matter. It concerns the most important contribution that cosmology has made to our understanding of the world.

The import of the second cosmological fallacy is that our received image of both nature and natural science is modeled on a historically parochial view of how nature works. Cosmology has long since denied us any entitlement to such parochialism. We nevertheless remain reluctant to give it up.

There is no invariant or quintessential scientific method. Our views of the practice of science develop together with the content of our scientific ideas. The discovery that the universe has a history, and so therefore must everything within it be historical, has implications for the practice of science. We have so far failed to acknowledge them.

It is not just any history. It is a particular history. We already know enough about it to begin to form the idea that nature can exist in different states or wear different masks. Our prevailing conception and practice of scientific explanation take only one of these states for granted, and identify that state with the necessary and universal workings of nature. In so doing, however, they fail to take adequate account of what cosmology has already discovered to be the facts of the matter – at least if we interpret its findings undistracted by metaphysical prejudice.

Yet on any of the accounts of the origins of the present universe that now command authority, we have reason (although we have no direct evidence) to think that the workings and characteristics of nature were once very different from what they have since become. These views can be broadly grouped into two main families of ideas.

One family of ideas, predominant to this day, traces the origins of the observed universe to a singularity in the now conventional sense: an original state in which the energy density of nature reached infinite value. Another family of ideas, which the argument of this book accepts, follows the history of the present universe to an original state in which the energy density of nature was extreme but nevertheless finite. In this second family of ideas, the conjecture of an original state of extreme energy density is readily married to the further conjecture of a succession of universes.

What is striking is that on either of these two sets of conceptions, we have reason to suppose that the familiar divisions within the mature and evolving universe – the structural distinctions and rudimentary components of nature described at one level by the standard model of contemporary particle physics and at another level by the periodic table – may once not have existed. (Of course, the chemical description of nature, as summarized in Mendeleev's periodic table, is not fundamental. It is nevertheless connected through many intermediate links, such as the Dirac equation and the Pauli exclusion principle, to the fundamental description offered by particle physics. The idea of a permanent differentiated structure of natural phenomena is central to the dominant tradition of modern science. Darwinism and, more broadly, the earth and life sciences have barely made a dent in the ascendancy of this vision. The idea of an ahistorical differentiated structure does not live exclusively in the forms of science that explore fundamental levels of reality – particle physics first among

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

them – but also in the sciences – chemistry, for example – that address nature at less fundamental levels. Unless we are to subscribe to a radical reductionism, incompatible with the way in which the physical sciences have developed, there is no reason to disregard the less fundamental descriptions and to focus solely on the more fundamental ones. The notion of a timeless structure must be contested and overthrown at all levels: the less fundamental ones as well as the more fundamental ones. In the meantime, chemistry, like particle physics, continues to be a structural science rather than a historical one.)

It is not simply, on this line of reasoning, that other structural distinctions and rudimentary components marked nature in the very early universe. It is that the presentation of nature as a differentiated structure, and its working as an interaction of clearly distinct forces or fields, may then have failed to obtain. Such distinctions and interactions could not have existed under the conditions of the very early universe. If they existed at all in the circumstance of the original extremes, they would have had to have been radically different, and to have worked in a radically different way, from how they later came to be and to work. A premise of much established thinking in cosmology and physics is, nevertheless, that nature works always and everywhere as a structure of distinct parts (particles, fields, forces) interacting with one another in conformity to unchanging laws.

The criticism of the second cosmological fallacy has as its aim to explore this contradiction within our present beliefs about the history of the universe and to consider its implications for the practice of science as well as for the content of some of our most comprehensive scientific theories. The whole argument represents a natural-philosophical reflection on what it would mean to take altogether seriously the idea that has been central to cosmology ever since Lemaître's conjecture about the origins of the universe gained widespread acceptance.

In this reflection, I resort to a heuristic device. I imagine two states of nature and say nothing about the transition from one to the other. This contrast, in the terms in which I sketch it, far exceeds the authority of the evidence. Moreover, it is couched in terms that could not figure among the formulations of a developed scientific theory. Nevertheless, it serves a legitimate analytic purpose: the aim of exposing the logic of the idea that nothing in nature lasts forever. In particular, it makes this logic explicit in the context of the second family of beliefs to which I have just referred: those that presuppose extreme but not infinite values of the earliest states of affairs in the history of the universe. The strategy of the heuristic argument is to contrast only two states of nature and to suggest nothing about the transition from one to the other.

The stark simplifications and the metaphorical language to which I here appeal in no way undermine the usefulness of the argumentative device. The core point is that the research agenda and the way of thinking inspired by Lemaître's conjecture fail to be fully achieved if we content ourselves with the idea that the early universe had a different structure. The implication of the conception of the original state is that it had no structure at all, in the familiar sense of the concept of structure to which the scientific study of the mature universe has accustomed us. Because it had no such structure, it must also be supposed to have worked in a different way.

Moreover, the significance of the device is not limited to finitistic views of the original state: accounts of the original state of the universe that restrict all parameters to finite values. It is pertinent as well, albeit in a different way, to views that invoke a finitude-defying singularity. For, according to such views, there must also have been a moment when the distinctions and interactions of the mature universe did not yet exist. There must have been a transition and a transformation leading from the universe then to the universe later. Indeed, the transition and the transformation must have been all the more far-reaching if they accommodated, as they must have for such conceptions to make sense, a passage from the infinite to the finite.

In one state, nature appears and works as it does in the formed, cooled-down universe: the universe that we observe. Nature is divided up into discontinuous elementary components, the most basic of which are the particles, fields, and interactions studied by particle

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

physics. More generally, nature is constituted in this state by kinds of things or natural kinds: a fact that inspires the projects of classical ontology as well as of natural science. It is in this way that nature was seen in the tradition of Aristotle. It is likewise in this way that nature continues to be represented in the tradition that began with Galileo and Newton and continues to today.

Natural phenomena present themselves, according to this conception, within a limited range of parameters of energy and temperature. They display only modest degrees of freedom. The penumbra of the adjacent possible around each phenomenon – what it can become next, given what it is then – remains restricted or thin. The laws of nature – both the effective laws operating in particular domains, and the fundamental laws or principles cutting across domains – are clearly distinct from the phenomena that they govern. It is only a short step from these conceptions to the idea that changing states of affairs are governed by unchanging laws.

Nature, however, to follow the logic of this heuristic device, admittedly beyond the boundaries of the evidence before us but not contradictory to any of it, may also appear in another mode. It may have existed in this other way in the very early history of the present universe as well as at the beginnings or at the ends of other universes, if our universe was preceded by earlier ones. Nature may so appear again in its very late history. It may also from time to time present thus in particular regions, subject to extreme conditions. These local realities would then depart from the model of the workings of nature established in the cooled-down universe.

In this second state of the universe (the first, however, in the order of time), the structural distinctions among elementary constituents of nature have broken down or not yet taken shape. The parameters of temperature and energy are extreme but they are not infinite (as they are under the standard concept of a cosmological singularity). Consequently, no insuperable obstacle of principle exists to investigating and explaining them; it is not true that nature is open to our understanding only in its first state but not in its second.

Much higher degrees of freedom are excited than we observe in the cooled-down universe, and the penumbra of the adjacent possible around each phenomenon now becomes thick and rich. It does so whether we account for this wealth of transformative opportunity in the language of either causal or statistical determination. The laws – at least the effective laws applicable to particular domains – cease to be readily distinguishable from the states of affairs that they govern. If the phenomena change, the laws change coevally with them. This last characteristic of the second state of nature is intimately related to all the other traits: to the absence of clear and stable structural divisions (and thus of distinct domains to which different sets of effective laws would apply); to the extreme though finite physical parameters; and to the enhanced degrees of freedom enjoyed by the phenomena – the range of other phenomena that the existing phenomena can become and the facility with which they can turn into them.

The second cosmological fallacy is the disposition to take account of only the first state of nature while disregarding the second, and to do so in our methods as well as in our theories. When we succumb to this fallacy, our conception of how to practice science, as well as our view of the workings of nature, allows itself to be shaped by an intellectual engagement with only one set of the variations of nature. It becomes in a sense the science of a special case. It consequently remains limited in the reach of its insight even into that special case. The deepest enigmas of nature escape it.

It is not just the Newtonian paradigm that takes this path. It is an entire approach to science that has been shaped by the assumption that the first of these two states of nature (the second in the order of time) represents the ultimate and constant character of reality. In developing and supporting the idea that the universe has a history, cosmology, however, has already given us grounds to reject this assumption as false. On one interpretation of its findings (for which we argue in this book), everything is emergent – everything comes and goes – except time.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

The emergence of everything except time is one of the ways in which the first state of nature ceases to represent the essential and enduring character of reality. It is not the only way. Every version of the now standard account of the origins of the present universe suggests that nature at the earliest moments in the formation of the present, observed universe may have displayed traits very different from those that it later came to exhibit as it cooled down and assumed the structured form in which we now observe it.

It is simply that under many of the most influential cosmological theories – those that appeal to the idea of an initial singularity – the alternative traits of nature remain hidden under the veil of the infinite. The state that these theories purport to describe is one in which the parameters of the phenomena had infinite values. To ascribe infinite values to them is to place them effectively beyond the reach of inquiry and understanding: the ultimate secrets of universal history would remain sealed behind a door that we could never open. The result would be – indeed, it has been – to allow us to treat the variations and workings of nature as we encounter them this side of that door as if they were its permanent traits. It would also be to regard the practice of science that relies on this assumption as what science must always be.

If nature wears multiple disguises – the states through which it passes – a science that presupposes a stable structure of ultimate constituents of nature – the structure represented at one scale by the standard model of particle physics and at another scale by the periodic table – and a framework of immutable laws or symmetries clearly distinct from the phenomena that they govern cannot be more than the science of a special case, even if a special case of broad and enduring application. Such a science – the science that we in fact have – will be bereft of the cosmological equivalent of the physics of phase transitions: an account of the transitions from one state of nature to another.

Unlike the physics of phase transitions, such an account is universal rather than local. Unlike the physics of phase transitions, it requires a style of scientific explanation that dispenses with both the idea of a framework of immutable laws of nature and the picture of nature as a differentiated structure, made up of distinct elementary constituents – forces, fields, and particles – interacting with one another in conformity to such laws.

* * *

The two cosmological fallacies are closely connected. They reinforce each other. They make each other seem to be unavoidable conceptions – indispensable to the practice of scientific inquiry – rather than the contestable options that they in fact are.

The second cosmological fallacy limits our understanding of the variations of nature. In so doing, it makes the cosmological use of the Newtonian paradigm seem less troubling than it would otherwise be. It fails to solve the problem of the breakdown, in a cosmological context, of any distinction between initial conditions and a local configuration space of law-governed phenomena. Similarly, it does nothing to show how we can be justified in using the Newtonian paradigm in a setting in which we have no hope of observing or preparing copies of the explained phenomena. Nevertheless, the second cosmological fallacy represents nature as working always and everywhere in the way in which the Newtonian paradigm supposes it to work: by the conformity of distinct elements or phenomena, within a differentiated structure, to changeless laws.

The first cosmological fallacy presupposes a view of the workings of nature that makes any other conception of how nature works seem to be incompatible with the requirements of science. All the better then if nature can provide us with an excuse for the limitations of our insight by taking refuge in an exceptional condition that, because it has infinite parameters, is forever barred to investigation and understanding. It is for this reason that the conventional idea of the cosmological singularity helps makes the universalization of the Newtonian paradigm seem legitimate. By associating the finite with the workings of nature in the cooled-down state of the universe and any other variant of nature with the impenetrable infinite, it lends appeal to the second cosmological fallacy.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

Despite their reciprocal connections, the two cosmological fallacies have different characters and consequences. The second is more fundamental, and more far-reaching in its implications, than the first.

The first cosmological fallacy commits a mistake of method, with empirical assumptions and implications. The second cosmological fallacy amounts to a mistake about the facts of the matter, with wide consequence for the practice of science. The matter that it mistakes is the most important in science: the nature and history of the universe.

The argument against the first cosmological fallacy ends in a negative claim: the claim that we are not entitled to apply to the whole world the methods and habits of mind that modern science has applied to parts of the world. This negative claim in turn evokes the need for a way of thinking different from the one that the Newtonian practice expresses.

The argument against the second cosmological fallacy results in a positive claim: the claim that there is already more – implied if not shown – in what science has discovered about the universe than our established natural philosophy – the lens under which we read these discoveries – is willing to countenance. It suggests that this something more is baffling but in principle not inscrutable and that our understandings have not yet caught up to our findings. It inspires the need for a practice of science that can persevere in the endeavor of scientific inquiry even when the two features of nature that have seemed most indispensable to science are missing: the presence of distinct and constant elements or types and their interaction according to law-like regularities.

The arguments against the two cosmological fallacies require us to think historically about nature and its laws. As a result, they force us to confront what we here call the conundrum of the meta-laws. If the laws of nature have a history inseparable from the history of nature, it seems unacceptable to say either that their history is itself law-governed or that it is not. If the history of the laws of nature is law-governed, we seem to have rescued part of the standard view of science only by equivocating about the reality of time and by separating the content of the laws from the vicissitudes of the phenomena. If their history is not law-governed, it appears to lack an explanation, in violation of the principle of sufficient reason. Moreover, our causal explanations, relying as they do on the picture of a law-governed world, are rendered insecure. They will remain insecure until we change our understanding of the relation between causal connections and laws of nature.

The meta-laws conundrum is central to the agenda of cosmology. The solution to this conundrum bears on the meaning of every proposition within natural science. Cosmology is not an afterthought to physics. It is the part of natural science that has the most general implications for all the other parts.

## CAUSALITY WITHOUT LAWS

The three central claims of this book (about the world, time, and mathematics) and the argument against the two cosmological fallacies cannot be advanced without revising our view of causality and of its relation to laws of nature.

The approach to causation that has been predominant for several centuries rests on two pillars. The first pillar is the notion of causal links as mental constructs rather than as real connections in nature. The second pillar is the principle that causal explanations presuppose laws of nature: the laws serve as the warrants justifying causal explanations. We cannot have the latter without invoking the former.

That we should understand causation as a device of the mind – a requirement of the way in which we cope with the world and seek to understand it – rather than as a description of the workings of nature has been the prevailing view in philosophy since Hume and Kant. According to this view, causality is an indispensable habit of the mind, a requirement of our efforts to make sense of reality, an unavoidable simplification, a proxy for ultimate truths about nature that are forever denied us. So long as the inquiries and actions that we undertake under the aegis of the idea of causality produce acceptable results, either as theory or as practice, we have no reason to rebel.

One of the many benefits that this view of causation as mental construct renders to the ruling ideas about nature and science is to disguise or muffle the disharmony between causal connections among parts of nature and relationships among mathematical propositions. Nature, it is believed, works according to laws that are written in the language of mathematics. But how can there be such a comprehensive consonance between nature and mathematics if causal relations imply time (as effects succeed their causes) whereas mathematical relations are timeless (as the conclusions of a mathematical inference are conceptually simultaneous with its premises or points of departure or, rather, have nothing to do with the passage of time)? By treating causality as a necessary projection of the mind onto the workings of nature, which we would otherwise be unable to decipher and which we can grasp only under the constraints of human understanding, we make the paradox of the application of the timeless to the time-bound seem less troubling.

That causal explanations depend on an appeal, however tacit, to laws as well as to symmetries and constants of nature is a proposition that may seem all but self-evident. If causality has a clear and constant meaning, its proper usage appears to imply an appeal to regularities of nature. These regularities are laws, symmetries, and constants. However, it is laws, rather than symmetries and constants, that are easier to enlist, and have been most commonly enlisted, in the effort to explain why or how the same effects follow in similar circumstances from the same causes. Under this view, the laws of nature not only account for recurrent causal connections, they also establish which circumstances count as similar.

Causality without laws would seem to be a senseless notion: what would make the effect follow the cause? Without laws, relations of cause and effect would, according to this widespread conception, be arbitrary – mere coincidences – or express something different from what they seem to reveal. For example, they might describe relationships of reciprocal implication, better represented in the language of mathematics than in the vocabulary of cause and effect.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

The view of causality as mental construct is not, strictly speaking, inseparable from the thesis that causal explanations presuppose laws of nature. Nevertheless the two ideas reinforce each other. Each makes the other look yet more natural. If causality represents an enabling condition of our ability to reason about reality, then we can easily extend this supposedly indispensable syntax of concepts to include the partnership between causal accounts and law-like explanations. If causal explanations rely, implicitly or explicitly, on an appeal to regularities, especially laws of nature, then we can have more confidence that whatever the limits on our power to grasp "things in themselves" may be, we can at least bring order and clarity to our practices of inquiry, and hope to distinguish justified from unjustified beliefs about the workings of nature.

We do better to destroy these twin bases of the modern view of causation and of laws, and to think in another way. A different conception fits better with the ideas of the singular existence of the universe, of the inclusive reality of time, and of the selective realism of mathematics that we here develop and defend. It also conforms more closely to the empirical and experimental spirit of science. Causal connections, according to this alternative view, form a real feature of nature. They are not just an indispensable invention or projection of the mind.

Because they are real features of nature, they can take as many different forms as nature takes in the course of the history of the universe. Whether causal connections are always law-like is not a matter that we can determine by investigating the logic of our conceptual categories or the implications of our habits of mind. It is something that we can clarify only by finding out how nature in fact works, not universally and once and for all but rather variably, over time. It depends on facts of the matter about nature, not just on facts of the matter about human understanding.

If change changes, if the forms of connection and transformation evolve in the course of the history of the universe together with the states of affairs, then the real causal connections that bind nature together and that we describe in our theories may also undergo transformation.

* * *

In the most rudimentary sense, a causal connection is the influence that a state of affairs exercises over what follows it. The key presupposition of causality is therefore not the recurrence of the same connections: their law-like form. It is time. If time is not real, causality, understood in this way, cannot be real. It must be assimilated, or reduced, to something else: for example, to relationships of reciprocal implication, such as those that mathematics and logic represent.

Causal relations usually connect recurrent phenomena. Such will ordinarily be the case in what I earlier called the first state of the universe (which is the second state in order of time): the state in which a fixed structure of distinct elements of nature (as described by particle physics and by the periodic table) has taken shape, the laws or regularities of nature can be clearly distinguished from the phenomena that they govern, and there are tight constraints on the change from one state of affairs into another.

However, it may also happen that phenomena have not yet become, or no longer are, recurrent, if only because no structure of distinct elements or parts of nature has been established or maintained, the laws of nature are not yet, or have ceased to be, distinguishable from states of affairs, and the range of transformative opportunity – for the change of some states of affairs into others – remains ample. In such a circumstance – what I called earlier the second state of nature (but the first in the order of time, as in the early history of the present universe) – there can be causality without laws.

Causality will then continue to describe real relations in the one real, time-haunted world. However, there will not then be the element of recurrence or repetition enabling us, in similar circumstances, to attribute the same effects to the same causes. In this sense, the world will then be lawless.

The character of causation in each of these variants of nature is not a subject separate from scientific inquiry; it forms part of that inquiry. A theory in physics, cosmology, or any other branch of science is, among other things, an account of the real workings and changes of nature. We may, however, seek to develop a view of the similarities and differences among such causal connections: of what they are and of how they change.

Such a view will belong as much to natural science as to natural philosophy, and serve as an example of the porousness of the boundaries between them. Recognition of the real, rather than ideal, character of causal connections makes it possible to affirm their mutability and variety.

Under this conception, the character of causality cannot be uniform for the reason that everything in nature changes over time, including the forms of connection and of change. However, although change changes, it changes on the basis of what it was before. One state of affairs influences the next one. One way in which a state of affairs shapes its sequel influences a subsequent way in which it exerts this power over its aftermath. We should thus expect that despite the absence of a single form and meaning of causation there will be a substantial overlap among the forms and meanings of causal connection over time, the time of the history of an evolving universe. The common thread will be influence upon succession: causation is always about how every state of affairs in nature influences the states of affairs that succeed it in time.

* * *

"In time" is the decisive qualification: a universe in which causal connections form no part of nature (because they are mere constructions of the human intellect) is one in which time plays a secondary or epiphenomenal role. In such a universe there may be time-reversible laws of nature, as in Newton's mechanics. Reversibility of the laws diminishes the reality of time. Or there may be a timeless relational grid, as in Leibniz. The existence of a grid of that kind solves the contradiction between time-bound causality and time-denying logical or mathematical implication by reducing the former to the latter. Or the view may be offered that an appearance of causal succession merely disguises the workings of some other providential force coordinating events in nature and producing the false impression of causal connection. Such was the doctrine of the occasionalists, like Malebranche. These positions in the natural philosophy of the seventeenth and eighteenth centuries may seem more or less quaint until we realize that they remain alive in other less evident and all the more dangerous contemporary counterparts.

The preceding contrasts show that the reality of causal connection is closely or internally related to the reality of time. This relation has at least three aspects. First and most fundamentally, causation takes place in time and implies the reality of time. Second, time would not be inclusively real if causal connections simply enacted timeless laws of nature. From the idea of causal connection as such an enactment, it is only one step to the notion of time-reversible laws of nature (as in Newtonian mechanics). Third, the variety and mutability of causal connections – properties that they can meaningfully possess only if they are realities of nature rather than simply constructions of the mind – help us better understand what is implied in the claim that time is real.

That everything in nature can change – the kinds of things that there are as well as the ways in which they change – means that nothing stands outside time. It also modifies our understanding of what time is: part of what is at stake in the thesis of the reality and inclusiveness of time is that no absolute framework, whether of space or of laws or of mathematical truths and relations, envelops time. It is time, on this view, that envelops everything else. It is the only feature of nature that enjoys absolutely the attribute of non-emergence.

On this account, the long-held conventional view of the relation between causal connections and laws of nature is turned upside down. It is the causal connections, not the laws of nature, that are primitive and fundamental, though also time-bound, diverse, and mutable. By the laws of nature, we designate a feature that causal connections sometimes

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

fail to possess: that they recur because they bind together recurrent phenomena.

In the mature, cooled-down universe, most natural phenomena possess this feature. Suppose, however, that we take the long, cosmological view, especially when we prefer the idea of a succession of universes, or of states or phases of the universe, to the idea of a plurality of universes, and reject the notion that the universe began in an infinite initial singularity. The way is then open to think that causal connections may at times have failed to work as recurrent connections among recurrent phenomena. They may have failed to exhibit the feature of recurrence in the early universe: the universe before (or after) a discriminate structure emerged and laws became distinguishable from states of affairs. They may again fail to exhibit that feature later on, in extreme states of nature during the evolution of the cooled-down universe.

In this conception, the laws, like the bonds of causality, represent real features of the workings of nature. They are no mere heuristic devices. Theirs, however, is a derivative reality by contrast to the primitive and fundamental reality of causal connections. The invocation of laws describes a special case – the standard case in the mature universe. By using the vocabulary of laws we allude, as if by shorthand, to defining features of this standard case: regularity in the ties among repetitious phenomena. It is, in more senses than one, the inverse of the now conventional account of the relation between causal connections and laws of nature. In that account, it is the causal connections that are derivative from the laws of nature, and affirmed only for the convenience of human understanding. If my argument is correct, we should invert this line of reasoning.

This inverted view has implications for the conundrum of the meta-laws: the problem of how to think about change of the laws of nature, given that either of two apparent solutions to this problem seems unacceptable. One of these solutions appeals to the idea of higher-order laws governing change of the laws. It triggers an infinite regress and circumscribes, unjustifiably, the inclusive reality of time. The other solution dispenses with the idea of higher-order laws. It makes the change of the laws seem to be uncaused, if indeed causation presupposes the operation of laws.

In discussing the conundrum of the meta-laws I suggested a response to the conundrum: the co-evolution of the laws and of the phenomena, an idea familiar in the life sciences as well as in social and historical study. This idea, however, remains incomplete and unnecessarily baffling if not complemented by the idea of the primitive reality of causal connections.

The idea of the co-evolution of laws and phenomena makes sense if, and only if, causal connections are real in nature. Because they are real, and imply time, indeed in a sense embody time, they can change over time. If causal connections were only mental constructs, to say that they change would be indistinguishable from saying that our ideas about them change. We would have no basis on which, and even no vocabulary with which, to distinguish change in theories about causal connections from change in such connections.

The idea of the reality of causal connections remains unfinished and enigmatic so long as it fails to be developed into a view of how, in the course of the history of the universe, causation acquires a law-like form. Such a view leads into an account of how the laws may change as the phenomena and their connections change.

Thus, it is a mistake to regard the idea of the co-evolution of laws and phenomena and the idea of the real and primitive character of causal connections as two separate conceptions, much less as rival ones. Rather they represent two aspects of the same approach. Together, they suggest the beginning of a solution to the conundrum of the meta-laws. They bring greater clarity and support to the central theses of this book: that there is only one real universe, that time is real and inclusive, and that mathematics gains its power by exploring a counterfeit version of the world, bereft of time and particularity. These ideas do their work at the cost of attacking the foundations on which much of our thinking about causes and laws has wrongly come to rest.

That the laws of nature supervene on causal connections, which are primitive in nature, is a view diametrically opposed to the conventional conception, according to which causal connections are mere instances of the laws of nature.* Causal connections regularly assume law-like form in the observed, cooled-down universe. There may, however, be states of natural reality in which they exhibit no such form. Such states (by inference from current standard cosmological ideas) may have played a central role in the formative moments of the present universe as well as in extreme conditions (such as those that prevail in the interior of black holes) occurring in its subsequent history.

That causality can exist without laws is a proposition that may seem paradoxical to the point of absurdity when entertained in the context of physics. Yet it has become a commonplace, though an inadequately explained one, in the life and earth sciences as well as in the study of society and history.

In Chapter 2, I discuss how this problem has been expressed in the history of social theory and of social science. Those who insist on the vital influence of formative institutional and ideological structures in society and of structural discontinuity in history are, for the most part, no longer able to believe in laws of historical change, driving forward the succession of such structures. They have, for example, largely abandoned explanatory practices, like the one Karl Marx embraced, that represent history as a law-like progression of indivisible

* It is also to be distinguished from views holding that the empirical discoveries of science are best understood without any reliance whatsoever on the idea of laws of nature in any state of the universe. See, notably, Bas C. van Fraassen, *Laws and Symmetry*, 1989, proposing that symmetries rather than laws deserve to be placed at the center of our understanding of scientific inquiry. In this argument I take invariant symmetries, just as I take laws of nature, to be a mode of causation rather than its basis. They characterize the workings of nature over much of the history of the universe. They need not characterize these workings always and everywhere. I focus on the relation of causes to laws rather than to symmetries because of the central role that the idea of timeless laws of nature has played in the development of the traditions that we here oppose. Regularities in the workings of nature are laws, symmetries, or constants. A symmetry may be defined informally as a transformation that leaves all relevant structure intact. Relevance is determined with respect to a theoretically chosen and interpreted context. The concept of symmetry is intimately related to the idea of invariance. Thus, symmetries, if invariance constitutes part of their nature, impose a restraint on the inclusive reality of time, as do immutable laws.

institutional regimes: the modes of production in his social theory. The task then becomes to do justice to causal influence and constraint in the succession of such regimes without appealing to unbelievable laws of history. This problem is analogous, in some ways but not in others, to the conundrum of causality without laws in cosmology and physics.

Change changes. That it changes is much of what the thesis of the inclusive reality of time means. The transformation of transformation implies that the laws of nature are in principle mutable. It also implies that the way in which a prior state of affairs can influence a later state of affairs, when causality exists without laws, can also change.

Causation works with what exists at any given time, including the established forms of change. It does not work by selecting from a range of states of affairs marked as possible according to the criterion of some abstraction from nature, such as the criterion of the varieties of phenomenal connection that we are able to represent mathematically. Nor does it operate by returning to some no longer existing form of connection, unless that prior form of connection retains a vestigial presence in the universe that now exists; otherwise, the recurrence would represent the temporal equivalent to action at a distance.

Wherever, as in most of the observed universe, there exists a differentiated structure, a clear distinction between states of affairs and laws of nature, and tight constraints on what can happen next, the change of change will be rare. It will take the form of the appearance of emergent phenomena, with new properties, displaying new regularities, or governed by new laws. Such is the case with the phenomena studied by the earth and life sciences, and then again with those realities that we address when we try, through the study of mind, society, and history, to understand ourselves. Complexity may expand the range of the adjacent possible – of the *theres* that nature can reach from any given *here*. In so doing, it creates a basis for emergent phenomena, exhibiting novel regularities.

Suppose nature can also exist in another form, the second state evoked in my discussion of the second cosmological fallacy, in which there is no differentiated structure and no clear contrast between laws

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

and phenomena, in which many degrees of freedom are excited, and in which there persists ample transformative opportunity. In such a state, the restraints on the change of change will be weakened. Degrees of freedom, the adjacent possible, and emergent phenomena and properties will no longer be concepts that can be clearly distinguished when applied to such a presentation of nature.

We are accustomed, by the dominant tradition of physics, established as the supreme model of successful science, to regard historical explanation as ancillary to structural explanation. On the view that we here defend, this hierarchy must be reversed: structure results from history. Historical explanation is, thus, more fundamental than structural explanation. Cosmology affirms its ambition to be the most comprehensive natural science when it understands itself as a historical science first, and as a structural science only second.

The primacy of historical over structural explanation should give no offense to science, so long as we qualify the demand for causal explanation of everything in two ways (neither of which would be acceptable to those who espouse the metaphysical rationalism of Leibniz's principle of sufficient reason). The first qualification is that we allow a historical explanation to count as a causal account in cosmology and physics as in other branches of sciences, indeed as the most characteristic form of causal explanation when the subject matter of science becomes the whole universe. Under a historical view, a state of affairs is the way it is because of the influence of an earlier state of affairs, not because it conforms to timeless and invariant regularities. We shall not always be able to account for the influence of the earlier on the later by invoking such regularities. The second qualification is that we be willing to pay the price of a practice of historical explanation that is not subordinate to structural explanation.

This price has, in turn, two parts. The first part is that there is no absolute beginning. Time, we argue in this book, is not emergent. At any given moment in the history of science, our ability to draw inferences, supported by observation, is limited. Moreover, even if it were unlimited, we could not peer into the beginning of time; on this

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

account, time has no beginning. Thus, historical explanation is by its nature incomplete.

The second part of the price is that change in how change occurs, as described by a historical science, has an ineradicable matter-of-fact-ness or facticity. We can increase the extent to which we are able to make sense of the transformation of transformation. At the end of the day, however, nature will always be found to have an irreducible factitious element: it is what it is. If it were not what it is, but rather the consequence of some mode of rational necessity, history would once again be subordinate to structure.

We can attenuate such just-so-ness. We cannot abolish it. Examples of how we can attenuate it are the proposals that we make later in this book for the resolution of the dilemma of the meta-laws in cosmology, conceived as a historical science: we have reason to resist accepting either that change of laws of nature is governed by higher-order laws or that it is not.

* * *

These propositions require us to believe that the workings of nature are not necessary, even though they are causally determined. There is no univocal, unambiguous notion of necessity in science. Necessity designates the limit of the least mutable realities that are represented in a given set of ideas: what is necessary is whatever, according to that way of thinking, could least be other than it is.

In the tradition of physics that began with Galileo and Newton, the content of this limiting ideal of necessity is given by the convergence of three commitments.

The first commitment is to what we call the Newtonian paradigm: the extrapolation to the whole universe of an explanatory strategy, distinguished by the contrast between initial conditions and timeless laws applying within a configuration space demarcated by stipulated initial conditions. This procedure is legitimate only when applied to parts of the universe. To repudiate its cosmological application was the aim of my argument against the first cosmological fallacy.

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

The second commitment is to the premise that the characteristic form of the observed, cooled-down universe, with its stable, differentiated structure, its apparent contrast between laws and phenomena, and its severe constraint on degrees of freedom, on the range of the adjacent possible, and on the facility for the appearance of emergent phenomena and properties, that is to say, of the new, is the only form of nature. To reject this temporal generalization of the form that nature takes in the cooled-down universe was the purpose of my criticism of the second cosmological fallacy.

The third commitment is to the sovereignty of mathematics over physics. On the view presupposed by that commitment, what is physically realized is what can be mathematically represented and justified. Mathematics stands to physics as both oracle and prophet, divining the ultimate nature of reality. Given the non-temporal and ahistorical character of the relations among mathematical propositions, this commitment is intimately related to the assumption of the immutability of the laws of nature and to the invariance of its symmetries, expressed as mathematical equations. One form of this ambition is to conceive the universe as isomorphic to a mathematical construction or even as a mathematical structure. Another form is to infer the laws and symmetries of nature from the most consistent and comprehensive mathematical ideas. To contest the third commitment is the goal of my discussion of mathematics in Chapter 6.

Neither any law or symmetry of nature, however fundamental it may appear to be, nor any working of causality, in the absence of such laws and symmetries, is necessary if by necessity we mean an idea of necessary realities and relations that is defined by the coexistence of these three commitments.

It does not follow, however, that the meaning of the thesis of causality without laws is to affirm the radical contingency of the way in which nature, at any given time, works. Radical contingency is a metaphysical, not a scientific, idea. Its function is to express a disappointment: that we cannot infer the way things are from the imperatives of reason (in the spirit of Leibniz's principle of sufficient reason). Its

https://doi.org/10.1017/CBO9781139696487.003 Published online by Cambridge University Press

invocation betrays bad faith or confusion: a surreptitious genuflection to rationalist metaphysics by those who pride themselves on having cast off its shackles. It is an homage that has often had an ulterior religious, moral, or political motive.

The way things are is, for science, just what they are. The subordination of structure to history ensures the defeat of the rationalizing metaphysical project that the dominant tradition in physics has patiently served. It has served this project in the conviction that in so doing it would be able to wed mathematics, and serve itself. As dowry, it received from mathematics a poisoned gift: the means with which to explain temporal events by timeless laws.

Structure results from history. The combination of fundamental historical explanation with derivative structural explanation is the basis of science.

It falls to science to make sense of how and why the workings of nature are what they are. To guard against illusion, it must do so, however, without taking the why part of this endeavor as an invitation to infer natural reality from rational necessity. The universe is more neutrally described as factitious than as radically contingent: its most important attribute is that it is what it is rather than something else. It is what it is because it was what it was.