Chapter 1

ULTIMATE EXPLANATIONS

1. TO UNDERSTAND UNDERSTANDING

A powerful, not fully comprehensible instinct to understand lies dreaming in us. We would like to have everything fully comprehended, explained, and proved. So that there should be nothing left without its cause, moreover a cause that would remove all the anxiety of doubt and all the question marks. The grander the matter, the more we want to have it explained, to eradicate even the slightest suspicion that things might be otherwise. This longing for "ultimate explanations" is not fully comprehensible in itself, and when we want to understand it a nagging question inevitably arises: what does "to understand" mean?

In trying to answer that question philosophers of science have used up forests of paper and a sea of ink. In the age when Positivism was the prevalent trend attempts were made to dismiss all questions for which there were no answers in sight within the grasp of direct experimental methods, and it was postulated that the task of science was not to explain but to describe. But already by the classical period in the development of the philosophy of science on the basis of the methodology proposed by the Vienna Circle (Wiener Kreis) the question was not whether science explained anything, but what scientific explanation *meant*. If we agree that a description is a set of statements giving information about something, while an explanation is, in the most general sense, also a set of statements, but one that shows a logical connection between those statements, it is evident that, on the one hand, there is no clear line of demarcation between description and explanation, but – on the other hand – that explanation is something more than description.

M. Heller, *Ultimate Explanations of the Universe*, DOI 10.1007/978-3-642-02103-9_1,

Even a cursory knowledge of any theory in physics is enough to realise that not only does it describe, but also that it explains. Or rather that it only describes once it has explained. The framework for every theory in physics is always a mathematical structure, usually an equation or set of equations, appropriately interpreted, that is with reference to the world. A mathematical structure is essentially a network of logical inferences appropriately encoded in symbols. To unlock and decode those inferences we have to resolve the mathematical structure computationally, usually by solving the equation or set of equations. The interpretation, in other words reference to the world, is not carried out directly, but by calculating the empirical predictions of the theory and comparing them with the results of what is actually observed. This procedure is in fact tantamount to inserting the results of experimental observation within the grid of inferences making up the theory's mathematical structure. Moreover, the world speaks only through the medium of mathematical structures. Experimental results are always expressed in terms of numbers, but these numbers are meaningless outside the structure containing them. The logic of the way the measuring apparatus works is essentially part of the logic of the mathematical structure serving as a framework for the particular physical theory. The structure of the given physical theory is as it were embodied in the construction of the apparatus.

A standard textbook of methodology envisages the distinction between description and explanation in the fact that a description is "a set of statements providing information on a particular aspect of reality without a clear reference of these statements to other statements," whereas an explanation is "a series of statements connected with each other by means of systematic proof."[1] If we accept these rather general expressions, then there is no such thing as a pure description in the theories of physics, a description is always an explanation. The same author writes that the distinction between description and explanation is of the same kind as between statements and proofs.[2] This formulation leads to the problem of ultimate explanations. In the process of proving something we cannot "go back to infinity," ultimately we have to adopt some axioms as the basis of our proof. Similarly in a description we have to start from a point of reference. Otherwise we risk moving back to infinity. So what does "the ultimate explanation" mean? We should find something that would be "an explanation for itself". Like God in Christian theology, who "Is because He Is" – the self-explanatory Absolute; His non-existence would mean a fundamental incongruity. But all the indications are that this Logic is not accessible to our reason, and if we want to rely on reason alone we must maintain a far-reaching respect with regard to theological rationale.

Obviousness, in the sense of something that is self-explanatory, turned out to be an embarrassment to the history of science long ago – from the time of the

Ptolemy–Copernicus controversy to the achievements of quantum physics. Our sense of the self-evident has grown up in the course of our encounter with the macroscopic environment and invariably fails us whenever we have to go beyond the borders of that environment. "Infinitely small" and "infinitely large" worlds are completely different from the one to which our eyes have become accustomed.

There is one more possibility: something in the nature of "circular explanations" – a closed chain of inferences: the current conclusion becomes the reason for the statements of which it is an inference. There are many ideologies which employ this sort of intuition to build up philosophical visions, but until a logical model is created showing that such an approach is not self-contradictory, they will only be visions and ideologies. Self-referential methods are applied fairly often in logic and mathematics, e.g. in the proof of Gödel's famous theorem, but such methods are still a long way off from what we would be inclined to call the ultimate explanation.

All these misgivings are not powerful enough to fetter the belief which is not only rooted deep inside our personal expectations but also resolutely set on the horizon of the ambitious human enterprise called science – the belief that everything has its reason. The endeavour to reach that horizon is the driving force behind science.

2. THE TOTALITARIANISM OF THE METHOD

At first sight the mathematical and experimental method of the contemporary sciences looks highly ascetic. Its very origins involved the withdrawal from overly complex metaphysical issues and a limitation to the analysis of the simple facts of experimental data. This constraint of the field of interest immediately brought a remarkable effectiveness. More and more phenomena, further and further away from ordinary experience, were subjected to the mathematical and experimental method, but the method itself continued to be interpreted very ascetically, within the bounds of the measurable. This approach gave rise to Positivism: whatever was beyond experiment was not worthy of scientific attention at all. With time experimental accessibility turned into a criterion determining existence. In its most radical phase Positivism tended to adopt a view that all that could not be verified experimentally simply did not exist. It is not difficult to spot an idiosyncratic kind of methodological totalitarianism lurking in this attitude. The mathematical and experimental method simply does not tolerate competition: whatever resists its application is annihilated. In its Neo-Positivist version this totalitarianism was reduced to a claim that the bounds of rationality coincided with the boundary of the mathematical and experimental method. Whatever lay beyond the confines of this method was beyond the reach of rationality and was therefore irrational, in other words bereft of a sense.

The consequence of such an attitude should be the conviction that the ultimate explanation of the universe lies within the grasp of the mathematical and experimental method. For if there are no explanations other than those obtained on the grounds of this method, then the ultimate explanation – the furthest-reaching explanation – must lie within its bounds. However, such notions were not voiced openly in the Positivist and Neo-Positivist era, since they did not concur with the Positivist principle of economy. The postulate of ultimate explanations smacked of metaphysics, which had been relegated, not only from the realm of science, but also from areas connected with science. Today, after the demise of classical Positivism, this attitude survives only in a few of the more radical groups of analytical philosophers. Many scientists, liberated from the Positivist scientistic straitjacket, are succumbing to the natural instinct to search for ultimate explanations, but are setting about it as it were "on an extension" of the mathematical and experimental method, not really worried by the fact that at a certain point on this quest the boundary between physics and metaphysics must inevitably be crossed. In general it is claimed that the right place for such endeavours is the popular scientific literature meant for the general public, while in the professional publications on their research scientists tend to refrain from embarking on philosophical deliberation. This is only part of the truth, since apart from overt forays into philosophy there are also a variety of highways and byways on which philosophy may creep into scientific research. One of them is the pursuit of theories and models which offer a chance for an "ultimate explanation" in a version sensed, or even concocted, by the scientist in question. What is more, on closer scrutiny of the history of science it turns out that this is a strategy that worked well even in the heyday of Positivism.

The tendency to pursue "ultimate explanations" is inherent in the mathematical and experimental method in yet another way (and another sense). Whenever the scientist faces a challenging problem, the scientific method requires him never to give up, never seek an explanation outside the method. If we agree – at least on a working basis – to designate as the universe everything that is accessible to the mathematical and experimental method, then this methodological principle assumes the form of a postulate which in fact requires that *the universe be explained by the universe itself.* In this sense scientific explanations are "ultimate," since they do not admit of any other explanations except ones which are within the confines of the method.

However, we must emphasise that this postulate and the sense of "ultimacy" it implies have a purely methodological meaning, in other words they oblige the scientist to adopt an approach in his research *as if* other explanations were neither existent nor needed. It is a psychological fact that people who practise the scientific method for a long time develop a compulsive habit of endowing methodological rules with an ontological sense, in other words they become convinced that explanations which transcend the mathematical and experimental method are pseudo-explanations, since nothing exists beyond the reach of this method. This is evidently a path straight into the Positivist ideology, and if for any reason the scientist does not want to succumb to it, he has to "extend" the scientific method to cover all that he wishes to study. Today the latter tendency seems to have a goodly following of adherents. Many scientists are probably not at all aware of the need to distinguish between the "methodological order" and the "ontological order," and treat methodological rules as if they were ontological principles.

Let us return to the postulate to "explain the universe by means of the universe itself." The word "universe" in this expression, which sounds rather like a slogan, clearly indicates that the science in which the drive to seek "ultimate explanations" is the most manifest is cosmology – the science of the universe. On the one hand, cosmology, envisaged in a certain sense as dealing with a complete totality, has no chance at all of seeking any explanations beyond its own area of research; but on the other hand, for instance in its formulation of the question of the origins of cosmic evolution, it as it were imposes an extraneous perspective. Here the distinction between the "methodological order" and the "ontological order" turns out to be very useful. But no distinctions can remove the tension between the tendency to be rigorously economical in research methods, and the longing for full understanding. It is in the field of cosmology that the controversy over "ultimate explanations" is raging at its most excited and impassioned state.

3. MODELS

In 1983 Jim Hartle and Steven Hawking published a paper in which they proposed the later renowned model for "the quantum creation of the universe out of nothing."[3] Their principal aim was to unify the general theory of relativity, in other words Einstein's theory of gravitation, with quantum physics, making up a single, integrated theory of physics. In this paper they put forward an approximate scheme for the quantisation of gravitation and tried to show that within the framework of this scheme there existed a finite probability of the universe emerging in a certain state out of an "empty" state. They called this mechanism "the quantum creation of the universe out of nothing," which became the point of departure for many other papers and a sort of paradigmatic example of "the ultimate explanation in cosmology." We shall devote one of the following chapters to the Hartle-Hawking model.

One of Hawking's students, Wu Zhong Chao from China, was so fascinated by the Hartle-Hawking model, which Hawking and his students later continued to develop, that he dedicated a separate monograph to the subject, published in English in China.[4] In this book, particularly at the beginning of Chap. 3, Wu makes a number of comments of a methodological nature on the subject of "the ultimate explanation" in cosmology. They apply directly to Hawking's model, but in fact are more general in character and therefore deserve closer attention here.

Like all other theories in physics, cosmological theories must obey the same principles regarding proper method: above all they must be self-consistent, viz. not mutually contradictory from the logical point of view, and at least not contradictory with respect to the observed empirical facts. The former requirement must be kept rigorously, while adherence to the latter requirement allows of a certain degree of tolerance described at length in contemporary textbooks of the philosophy of science. The point is that sometimes it is better to have a theory which has problems with explaining certain "slight experimental discrepancies" than to have no theory at all. This was the situation in the late nineteenth century, for example, when it was already known that Newton's theory of gravitation failed to explain certain "small aberrations" in the orbit of Mercury (its perihelion motion), but the theory still continued to be used very successfully. As we know, it is not easy to codify all the rules of the methodology of science, but in the everyday situations of research common sense supported by tradition and experience tells the scientist the right way to proceed. Research in cosmology must obviously comply with these procedures.

But cosmology has its own specifics. Wu also requires cosmological theories to be self-contained. Let's explain what he means by this. Usually the mathematical backbone of any physical theory entails a differential equation or a set of differential equations. Not only is there a need for such an equation (or set of equations) to be solved, but also the initial or boundary conditions have to be determined (or "imposed by hand," as physicists say) for the selection of the right solution for the particular physical problem. For instance, if the equation is to describe the motion of a given body, then the initial conditions may be its position and velocity at the start of the motion. If the equation is to describe the gravitational field of a given star, then we may select the behaviour of that field at an appropriate distance away from the star as the boundary conditions for the right solution, e.g. we may assume that at infinity (viz. an appropriately long distance away) the field strength is negligibly small. In other words it is the researcher who selects the initial or boundary conditions on the basis of his understanding of the physical situation he intends to model.

We could, of course, do likewise in cosmology. A cosmological model is a set of differential equations, too, and when we select a particular solution to it we also have to decide on particular initial or boundary conditions. Usually we will be guided in our choice by the principle of simplicity, or we will try all the possibilities and adapt the equation to the experimental data ex post, or else (as sometimes happens) we build up a "philosophy" to the solution we have found. The point is that none of these options is applicable to the physical situation in cosmology. Initial or boundary conditions are extraneous to the model; they are something the physicist has to "insert manually" into the model. The universe is a physical system still undergoing a process of evolution which cosmology is successfully reconstructing, thus its initial or boundary conditions must have been fixed in one way or another. Only there was no hand to manually insert them into the world. Or – to put it more precisely from the methodological point of view – within the framework of the mathematical and experimental method we are not allowed to assume that such a hand existed. We have to do without its assistance.

In a nutshell, we should look for initial or boundary conditions outside the universe. But we may not speak of the universe not having an exterior; we should rather say that the concept of an exterior is meaningless with respect to the universe. Wu calls this logical loop the problem of the First Cause,[5] which has been the bane of cosmology ever since Newton's time. The theory of cosmology would be self-contained if it managed to disentangle itself from this problem. As we shall see in subsequent chapters, cosmologists have been searching for such a theory (or model) along various paths. For instance we may imagine a theory which would not require any initial or boundary conditions, or a model which automatically determined its own conditions. We shall also see that many authors have resorted to highly exotic ideas to make the world self-contained. This is undoubtedly a philosophical motive, but one that derives from the right methodological perception, and is today a very powerful trend in the thought on the universe and cosmology.

4. ANTHROPIC PRINCIPLES AND OTHER UNIVERSES

The goal of unity is firmly encoded in scientific method. Modern science started when giants like Copernicus, Galileo, Kepler and Newton managed to unite "earthly physics" and "celestial physics," in other words to show that the same laws of physics hold on Earth and in astronomy. For some time thereafter it seemed that the laws of mechanics discovered by Newton were the ultimate, "unified" theory governing everything. The discovery of electricity and magnetism finally swept this illusion away, but soon Maxwell showed that these two classes of phenomena could be put together in a single, mathematically very elegant theory of electromagnetism. Einstein was the first to come up with the idea that Maxwell's theory should be combined with the theory of gravitation, and devoted the rest of his life to the implementation of this idea. Today we know that Einstein's concept had no chance of success, since apart from electromagnetism and gravitation there are two other fundamental physical forces: the weak nuclear force (the lepton interaction), and the strong nuclear force (the hadron interaction). We now have an experimentally confirmed theory (the Weinberg-Salam theory) uniting the electromagnetic force with the weak nuclear force in a single interaction, known as the electroweak interaction. In principle we also know how to combine the strong nuclear force with this interaction. We have several scenarios for the unification and are only waiting for the experimental data which will select the right scenario. Only gravitation, by virtue of its different character, is defying the successful application of the unifying schema with respect to itself. No wonder that more and more tension is building up in the search for a quantum theory of gravitation – for it is almost certain that gravitation will have to be quantised before it can be unified with the other interactions. Virtually every new mathematical model, as a rule more sophisticated than its predecessors, at first generates enthusiasm and new hopes, but is soon relegated to the gallery of interesting but abortive constructions.

The multiplicity of unifying models put forward hitherto is splitting up physics into separate schools and trends rather than uniting it. This process of fragmentation has reached an apogee in the theory which probably the largest group of physicists regard as the most promising today – the M-theory, a development and generalisation of the superstring theory, which has become well-known outside physics. According to the good old quantum theory there should be one minimum energy state, that is the fundamental state. In M-theory there is "practically an infinite number" of fundamental states (estimated at even as many as 10^{500}). The problem is that it is the fundamental state that to a large extent determines the physics of the universe. What are we to do with such a vast number of fundamental states? Unless we reject all of the theory leading up to them, the only solution is to accept that there is a very large number of different universes – as many as there are possible fundamental states – each with its own, different physics. M-theory people speak of the "string landscape" of the various universes, and conduct research on it.

It was easier to accept such a situation psychologically, as the idea of many universes had been circulating for some time in discussions on certain issues in cosmology. It first appeared in connection with the anthropic principles, which in various ways formulated the observation that the existence of living organisms on at least one planet in the universe depended in a very sensitive manner on the universe's initial conditions and its other characteristic parameters. A slight change in any of these conditions or parameters in general gives rise to drastic changes in the universe's evolution, rendering the emergence of biological evolution impossible. For instance, a very small change in the universe's initial rate of expansion (either its slight acceleration or retardation) would preclude the emergence of carbon, the very cornerstone of organic chemistry. There are many such "coincidences."

What has made the universe "friendly to life"? This brings to mind the notion of a purposeful design of the universe. But such an idea is alien to the rule of "explaining the universe by means of the universe itself." To neutralise it, the following argument was employed: suppose there exists an infinite number of universes representing all the possible combinations of initial conditions and other parameters characteristic of the given universe. Only a very few of those universes are life-friendly, and we live in one of them, since we could not have come into existence in any other. Some of those taking part in the discussion immediately acknowledged the hypothesis of many universes as more rational than the hypothesis of the existence of God, while others said that there was no need to "proliferate existences" if a single God was enough.

Regardless of these theological disputes, the idea of a multiverse (as it soon came to be called) launched into a life of its own. Soon fairly concrete cosmological models, e.g. inflationary models or certain unifying scenarios, started to identify mechanisms which could have produced either different universes, completely separated from ours, or regions in our own universe to which we shall never have perceptive access. At any rate, the multiverse trend became a reality. But was it still science? Can something which we will never be able to access by observation, even in principle, still be a subject for scientific study? Or maybe the scientific method was being transformed before our very eyes, and something that was not science before was now turning into science? Nonetheless, I think we should not be too hasty in undermining scientific method – that's right: in rashly undermining the scientific method, which is rightly regarded as the greatest achievement of science, and on which all the other achievements of science depend. Instead we should once more recall that the boundaries of rationality do not coincide with the bounds of scientific method, and therefore it is sometimes worthwhile to transcend those boundaries in order to be able to carry on a rational discourse "once on the other side." Although we should not expect any empirical solutions in that area, critical argumentation and rational appraisal will still be relevant there.

5. CREATION OF THE UNIVERSE

The discussions on the anthropic principles and the multiverse are as it were on an extension of scientific research. On the whole it is quite difficult to identify the point at which we cross the boundary between what may still be called a cosmological model, and what definitely belongs to speculation beyond that boundary. But we could be even more daring and locate our observation point well beyond the boundary of scientific method (yet still within the area of rationality), and once we are on the other side take a look how the mathematical and experimental method works within the area proper to it, and what happens to its explanations as it approaches the limits of its possibilities. The area "beyond" is very well-known in the history of human thought: it is the region inhabited by philosophical and theological concepts. It is vast and highly "speculative." To prevent us from losing our way on its tortuous paths, I shall limit the area by applying two restrictions: First, in practice I shall not go beyond the concept of creation as rooted in Judaeo-Christian thought. The concept of creation undoubtedly entails the ambition to explain, though in a theological sense. It is a theological concept, but has acquired numerous philosophical accounts (in the light of diverse philosophical systems), and it is chiefly the philosophical aspect of the idea which will be our subject of study. I shall touch on other philosophical concepts of ultimate explanations (or attempts to undermine them) only incidentally, more for the sake of a fuller picture of the philosophical ideas involved than of their analysis in depth. Secondly, out of all the versions and interpretations of the creation concept I shall only select ones which may be referred in one way or another to contemporary science, or those which, albeit historically distant from the present times, are still indispensable for the right understanding of such reference. This criterion is not so restrictive, as the history of ideas, in science as well as in philosophy, shows that ever since Christian Antiquity right through to the modern period, the mainstream thought on creation has been strictly linked genetically with the evolution of the ideas which led up to the development of the modern sciences. However, it is not my intention to compile a history of those genetic links, but to try to look at ultimate explanations from a different perspective than the one usually taken up on the grounds of physics and cosmology.

Is the philosophical and theological speculation located on a long extension of the investigations based on scientific theories and models? Perhaps the two are somehow mutually complementary to each other? Or perhaps – as some people claim – although apparently concerned with much the same thing, scientific inquiry and theological and philosophical deliberation are mutually untranslatable? Regardless of which of these possibilities (or maybe yet another one) is true, all of them are an expression of the same instinct ingrained in human rationality: to leave no stone unturned in seeking for an answer to every valid question.