Contents

Bibliography

Acknowledgements

I am grateful to my friends and colleagues Sarah Fitz-Claridge, Alan Forrester, Herbert Freudenheim, David Johnson-Davies, Paul Tappenden and especially Elliot Temple and my copy-editor, Bob Davenport, for reading earlier drafts of this book and suggesting many corrections and improvements, and also to those who have read and helpfully commented on parts of it, namely Omri Ceren, Artur Ekert, Michael Golding, Alan Grafen, Ruti Regan, Simon Saunders and Lulie Tanett.

I also want to thank the illustrators Nick Lockwood, Tommy Robin and Lulie Tanett for translating explanations into images more accurately than I could have hoped for.

Introduction

Progress that is both rapid enough to be noticed and stable enough to continue over many generations has been achieved only once in the history of our species. It began at approximately the time of the scientific revolution, and is still under way. It has included improvements not only in scientific understanding, but also in technology, political institutions, moral values, art, and every aspect of human welfare.

Whenever there has been progress, there have been influential thinkers who denied that it was genuine, that it was desirable, or even that the concept was meaningful. They should have known better. There is indeed an objective difference between a false explanation and a true one, between chronic failure to solve a problem and solving it, and also between wrong and right, ugly and beautiful, suffering and its alleviation – and thus between stagnation and progress in the fullest sense.

In this book I argue that all progress, both theoretical and practical, has resulted from a single human activity: the quest for what I call good explanations. Though this quest is uniquely human, its effectiveness is also a fundamental fact about reality at the most impersonal, cosmic level – namely that it conforms to universal laws of nature that are indeed good explanations. This simple relationship between the cosmic and the human is a hint of a central role of *people* in the cosmic scheme of things.

Must progress come to an end – either in catastrophe or in some sort of completion – or is it unbounded? The answer is the latter. That unboundedness is the 'infinity' referred to in the title of this book. Explaining it, and the conditions under which progress can and cannot happen, entails a journey through virtually every fundamental field of science and philosophy. From each such field we learn that, although progress has no necessary end, it does have a necessary beginning: a cause, or an event with which it starts, or a necessary condition for it to take off and to thrive. Each of these beginnings is ‘the beginning of infinity’ as viewed from the perspective of that field. Many seem, superficially, to be unconnected. But they are all facets of a single attribute of reality, which I call *the* beginning of infinity.

1

The Reach of Explanations

> *Behind it all is surely an idea so simple, so beautiful, that when we grasp it – in a decade, a century, or a millennium – we will all say to each other, how could it have been otherwise?*
>
> John Archibald Wheeler, *Annals of the New York Academy of Sciences*, 480 (1986)

To unaided human eyes, the universe beyond our solar system looks like a few thousand glowing dots in the night sky, plus the faint, hazy streaks of the Milky Way. But if you ask an astronomer what is out there in reality, you will be told not about dots or streaks, but about *stars*: spheres of incandescent gas millions of kilometres in diameter and light years away from us. You will be told that the sun is a typical star, and looks different from the others only because we are much closer to it – though still some 150 million kilometres away. Yet, even at those unimaginable distances, we are confident that we know what makes stars shine: you will be told that they are powered by the nuclear energy released by *transmutation* – the conversion of one chemical element into another (mainly hydrogen into helium).

Some types of transmutation happen spontaneously on Earth, in the decay of radioactive elements. This was first demonstrated in 1901, by the physicists Frederick Soddy and Ernest Rutherford, but the concept of transmutation was ancient. Alchemists had dreamed for centuries of transmuting 'base metals', such as iron or lead, into gold. They never came close to understanding what it would take to achieve that, so they never did so. But scientists in the twentieth century did. And so do stars, when they explode as supernovae. Base metals can be transmuted into gold by stars, and by intelligent beings who understand the processes that power stars, but by nothing else in the universe.

As for the Milky Way, you will be told that, despite its insubstantial appearance, it is the most massive object that we can see with the naked eye: a *galaxy* that includes stars by the hundreds of billions, bound by their mutual gravitation across tens of thousands of light years. We are seeing it from the inside, because we are part of it. You will be told that, although our night sky appears serene and largely changeless, the universe is seething with violent activity. Even a typical star converts millions of tonnes of mass into energy every second, with each *gram* releasing as much energy as an atom bomb. You will be told that within the range of our best telescopes, which can see more galaxies than there are stars in our galaxy, there are several supernova explosions per second, each briefly brighter than all the other stars in its galaxy put together. We do not know where life and intelligence exist, if at all, outside our solar system, so we do not know how many of those explosions are horrendous tragedies. But we do know that a supernova devastates all the planets that may be orbiting it, wiping out all life that may exist there – including any intelligent beings, unless they have technology far superior to ours. Its neutrino radiation alone would kill a human at a range of billions of kilometres, even if that entire distance were filled with lead shielding. Yet we owe our existence to supernovae: they are the source, through transmutation, of most of the elements of which our bodies, and our planet, are composed.

There are phenomena that outshine supernovae. In March 2008 an X-ray telescope in Earth orbit detected an explosion of a type known as a 'gamma-ray burst', 7.5 billion light years away. That is

Either way, the discoverer of knowledge is its passive recipient, not its creator.

But, in reality, scientific theories are not 'derived' from anything. We do not read them in nature, nor does nature write them into us. They are guesses – bold conjectures. Human minds create them by rearranging, combining, altering and adding to existing ideas with the intention of improving upon them. We do not begin with 'white paper' at birth, but with inborn expectations and intentions and an innate ability to improve upon them using thought and experience. Experience is indeed essential to science, but its role is different from that supposed by empiricism. It is not the source from which theories are derived. Its main use is to choose between theories that have already been guessed. That is what 'learning from experience' is.

However, that was not properly understood until the mid twentieth century with the work of the philosopher Karl Popper. So historically it was empiricism that first provided a plausible defence for experimental science as we now know it. Empiricist philosophers criticized and rejected traditional approaches to knowledge such as deference to the authority of holy books and other ancient writings, as well as human authorities such as priests and academics, and belief in traditional lore, rules of thumb and hearsay. Empiricism also contradicted the opposing and surprisingly persistent idea that the senses are little more than sources of error to be ignored. And it was optimistic, being all about obtaining new knowledge, in contrast with the medieval fatalism that had expected everything important to be known already. Thus, despite being quite wrong about where scientific knowledge comes from, empiricism was a great step forward in both the philosophy and the history of science. Nevertheless, the question that sceptics (friendly and unfriendly) raised from the outset always remained: how can knowledge of what has *not* been experienced possibly be 'derived' from what *has*? What sort of thinking could possibly constitute a valid derivation of the one from the other? No one would expect to deduce the *geography* of Mars from a map of Earth, so why should we expect to be able to learn about *physics* on Mars from experiments done on Earth? Evidently, logical deduction alone would not do, because there is a logical gap: no amount of deduction applied to statements describing a set of experiences can reach a conclusion about anything other than those experiences.

The conventional wisdom was that the key is *repetition*: if one repeatedly has similar experiences under similar circumstances, then one is supposed to 'extrapolate' or 'generalize' that pattern and predict that it will continue. For instance, why do we expect the sun to rise tomorrow morning? Because in the past (so the argument goes) we have seen it do so whenever we have looked at the morning sky. From this we supposedly 'derive' the theory that under similar circumstances we shall always have that experience, or that we probably shall. On each occasion when that prediction comes true, and provided that it never fails, the probability that it will always come true is supposed to increase. Thus one supposedly obtains ever more reliable knowledge of the future from the past, and of the general from the particular. That alleged process was called 'inductive inference' or 'induction', and the doctrine that scientific theories are obtained in that way is called *inductivism*. To bridge the logical gap, some inductivists imagine that there is a principle of nature – the 'principle of induction' – that makes inductive inferences likely to be true. 'The future will resemble the past' is one popular version of this, and one could add 'the distant resembles the near,' 'the unseen resembles the seen' and so on.

But no one has ever managed to formulate a 'principle of induction' that is usable in practice for obtaining scientific theories from experiences. Historically, criticism of inductivism has focused on that failure, and on the logical gap that cannot be bridged. But that lets inductivism off far too lightly. For it concedes inductivism's two most serious misconceptions.

First, inductivism purports to explain how science obtains *predictions about experiences*. But most of our theoretical knowledge simply does not take that form. Scientific explanations are about reality, most of which does not consist of anyone's experiences. Astrophysics is not primarily about *us* (what we shall see if we look at the sky), but about what stars are: their composition and what makes them shine, and how they formed, and the universal laws of physics under which that happened. Most of that has never been observed: no one has experienced a billion years, or a light year; no one could have been present at the Big Bang; no one will ever touch a law of physics – except in their minds, through theory. All our predictions of how things will *look* are deduced from such explanations of how things *are*. So inductivism fails even to address how we can know about stars and the universe, as distinct from just dots in the sky.

The second fundamental misconception in inductivism is that scientific theories predict that 'the future will resemble the past', and that 'the unseen resembles the seen' and so on. (Or that it 'probably' will.) But in reality the future is unlike the past, the unseen very different from the seen. Science often predicts – and brings about – phenomena spectacularly different from anything that has been experienced before. For millennia people dreamed about flying, but they experienced only falling. Then they discovered good explanatory theories about flying, and then they flew – in that order. Before 1945, no human being had ever observed a nuclear-fission (atomic-bomb) explosion; there may never have been one in the history of the universe. Yet the first such explosion, and the conditions under which it would occur, had been accurately predicted – but not from the assumption that the future would be like the past. Even sunrise – that favourite example of inductivists – is not always observed every twenty-four hours: when viewed from orbit it may happen every ninety minutes, or not at all. And that was known from theory long before anyone had ever orbited the Earth.

It is no defence of inductivism to point out that in all those cases the future still does 'resemble the past' in the sense that it obeys the same underlying laws of nature. For that is an empty statement: *any* purported law of nature – true or false – about the future and the past is a claim that they 'resemble' each other by both conforming to that law. So that version of the 'principle of induction' could not be used to derive any theory or prediction from experience or anything else.

Even in everyday life we are well aware that the future is unlike the past, and are selective about which aspects of our experience we expect to be repeated. Before the year 2000, I had experienced thousands of times that if a calendar was properly maintained (and used the standard Gregorian system), then it displayed a year number beginning with '19'. Yet at midnight on 31 December 1999 I expected to have the experience of seeing a '20' on every such calendar. I also expected that there would be a gap of 17,000 years before anyone experienced a '19' under those conditions again. Neither I nor anyone else had ever observed such a '20', nor such a gap, but our explanatory theories told us to expect them, and expect them we did.

As the ancient philosopher Heraclitus remarked, 'No man ever steps in the same river twice, for it is not the same river and he is not the same man.’ So, when we remember seeing sunrise ‘repeatedly’ under ‘the same’ circumstances, we are tacitly relying on explanatory theories to tell us which combinations of variables in our experience we should interpret as being ‘repeated’ phenomena in the underlying reality, and which are local or irrelevant. For instance, theories about geometry and optics tell us not to expect to see a sunrise on a cloudy day, even if a sunrise is really happening in the unobserved world behind the clouds. Only from those explanatory theories do we know that failing to see the sun on such days does not amount to an experience of its not rising. Similarly, theory tells us that if we see sunrise reflected in a mirror, or in a video or a virtual-reality game, that does not count as seeing it twice. Thus the very idea that an experience has been repeated is not itself a sensory experience, but a theory.

So much for inductivism. And since inductivism is false, empiricism must be as well. For if one cannot derive predictions from experience, one certainly cannot derive explanations. Discovering a new explanation is inherently an act of creativity. To interpret dots in the sky as white-hot, million-kilometre spheres, one must first have thought of the idea of such spheres. And then one must explain why they look small and cold and seem to move in lockstep around us and do not fall down. Such ideas do not create themselves, nor can they be mechanically derived from anything: they have to be guessed – after which they can be criticized and tested. To the extent that experiencing dots ‘writes’ something into our brains, it does not write explanations but only dots. Nor is nature a book: one could try to ‘read’ the dots in the sky for a lifetime – many lifetimes – without learning anything about what they really are.

Historically, that is exactly what happened. For millennia, most careful observers of the sky believed that the stars were lights that is essential for the initiation of unlimited knowledge growth – the beginning of infinity.

The quest for authority led empiricists to downplay and even stigmatize *conjecture*, the real source of all our theories. For if the senses were the only source of knowledge, then error (or at least avoidable error) could be caused only by adding to, subtracting from or misinterpreting what that source is saying. Thus empiricists came to believe that, in addition to rejecting ancient authority and tradition, scientists should suppress or ignore any *new* ideas they might have, except those that had been properly 'derived' from experience. As Arthur Conan Doyle's fictional detective Sherlock Holmes put it in the short story 'A Scandal in Bohemia', 'It is a capital mistake to theorize before one has data.'

But that was itself a capital mistake. We never know any data before interpreting it through theories. All observations are, as Popper put it, *theory-laden*,* and hence fallible, as all our theories are. Consider the nerve signals reaching our brains from our sense organs. Far from providing direct or untainted access to reality, even they themselves are never experienced for what they really are – namely crackles of electrical activity. Nor, for the most part, do we experience them as being *where* they really are – inside our brains. Instead, we place them in the reality beyond. We do not just see blue: we see a blue sky up there, far away. We do not just feel pain: we experience a headache, or a stomach ache. The brain attaches those interpretations – 'head', 'stomach' and 'up there' – to events that are in fact within the brain itself. Our sense organs themselves, and all the interpretations that we consciously and unconsciously attach to their outputs, are notoriously fallible – as witness the celestial-sphere theory, as well as every optical illusion and conjuring trick. So we perceive *nothing* as what it really is. It is all theoretical interpretation: conjecture.

Conan Doyle came much closer to the truth when, during 'The Boscombe Valley Mystery', he had Holmes remark that 'circumstantial evidence' (evidence about unwitnessed events) is 'a very tricky thing...It may seem to point very straight to one thing, but if you shift your own point of view a little, you may find it pointing in an equally uncompromising manner to something entirely different...There is nothing more deceptive than an obvious fact.' The same holds for scientific discovery. And that again raises the question: how do we know? If all our theories originate locally, as guesswork in our own minds, and can be tested only locally, by experience, how is it that they contain such extensive and accurate knowledge about the reality that we have never experienced?

I am not asking what authority scientific knowledge is derived from, or rests on. I mean, literally, by what process do ever truer and more detailed explanations about the world come to be represented physically in our brains? How do we come to know about the interactions of subatomic particles during transmutation at the centre of a distant star, when even the tiny trickle of light that reaches our instruments from the star was emitted by glowing gas at the star's surface, a million kilometres above where the transmutation is happening? Or about conditions in the fireball during the first few seconds after the Big Bang, which would instantly have destroyed any sentient being or scientific instrument? Or about the future, which we have no way of measuring at all? How is it that we can predict, with some non-negligible degree of confidence, whether a new design of microchip will work, or whether a new drug will cure a particular disease, even though they have never existed before?

For most of human history, we did not know how to do any of this. People were not designing microchips or medications or even the wheel. For thousands of generations, our ancestors looked up at the night sky and wondered what stars are – what they are made of, what makes them shine, what their relationship is with each other and with us – which was exactly the right thing to wonder about. And they were using eyes and brains anatomically indistinguishable from those of modern astronomers. But they discovered nothing about it. Much the same was true in every other field of knowledge. It was not for lack of trying, nor for lack of thinking. People observed the world. They tried to understand it – but almost entirely in vain. Occasionally they recognized simple patterns in the appearances. But when they tried to find out what was really there behind those appearances, they failed almost completely.

I expect that, like today, most people wondered about such things only occasionally – during breaks from addressing their more parochial concerns. But their parochial concerns *also* involved yearning to know – and not only out of pure curiosity. They wished they knew how to safeguard their food supply; how they could rest when tired without risking starvation; how they could be warmer, cooler, safer, in less pain – in every aspect of their lives, they wished they knew how to make progress. But, on the timescale of individual lifetimes, they almost never made any. Discoveries such as fire, clothing, stone tools, bronze, and so on, happened so rarely that from an individual's point of view the world never improved. Sometimes people even realized (with somewhat miraculous prescience) that making progress in practical ways would *depend* on progress in understanding puzzling phenomena in the sky. They even conjectured links between the two, such as myths, which they found compelling enough to dominate their lives – yet which still bore no resemblance to the truth. In short, they wanted to create knowledge, in order to make progress, but they did not know how.

This was the situation from our species' earliest prehistory, through the dawn of civilization, and through its imperceptibly slow increase in sophistication – with many reverses – until a few centuries ago. Then a powerful new mode of discovery and explanation emerged, which later became known as *science*. Its emergence is known as the *scientific revolution*, because it succeeded almost immediately in creating knowledge at a noticeable rate, which has increased ever since.

What had changed? What made science effective at understanding the physical world when all previous ways had failed? What were people now doing, for the first time, that made the difference? This question began to be asked as soon as science began to be successful, and there have been many conflicting answers, some containing truth. But none, in my view, has reached the heart of the matter. To explain my own answer, I have to give a little context first.

The scientific revolution was part of a wider intellectual revolution, the *Enlightenment*, which also brought progress in other fields, especially moral and political philosophy, and in the institutions of society. Unfortunately, the term 'the Enlightenment' is used by historians and philosophers to denote a variety of different trends, some of them violently opposed to each other. What I mean by it will emerge here as we go along. It is one of several aspects of 'the beginning of infinity', and is a theme of this book. But one thing that all conceptions of the Enlightenment agree on is that it was a *rebellion*, and specifically a rebellion against authority in regard to knowledge.

Rejecting authority in regard to knowledge was not just a matter of abstract analysis. It was a necessary condition for progress, because, before the Enlightenment, it was generally believed that everything important that was knowable had already been discovered, and was enshrined in authoritative sources such as ancient writings and traditional assumptions. Some of those sources did contain some genuine knowledge, but it was entrenched in the form of dogmas along with many falsehoods. So the situation was that all the sources from which it was generally believed knowledge came actually knew very little, and were mistaken about most of the things that they claimed to know. And therefore progress depended on learning how to reject their authority. This is why the Royal Society (one of the earliest scientific academies, founded in London in 1660) took as its motto 'Nullius in verba', which means something like 'Take no one's word for it.'

However, rebellion against authority cannot by itself be what made the difference. Authorities have been rejected many times in history, and only rarely has any lasting good come of it. The usual sequel has merely been that new authorities replaced the old. What was needed for the sustained, rapid growth of knowledge was a *tradition of criticism*. Before the Enlightenment, that was a very rare sort of tradition: usually the whole point of a tradition was to keep things the same.

Thus the Enlightenment was a revolution in how people sought knowledge: by trying *not* to rely on authority. That is the context in which empiricism – purporting to rely solely on the senses for knowledge – played such a salutary historical role, despite being fundamentally false and even authoritative in its conception of how science works.

One consequence of this tradition of criticism was the emergence of a methodological rule that a scientific theory must be *testable* (though this was not made explicit at first). That is to say, the theory must make predictions which, if the theory were nevertheless have horrified the early empiricists) that science cannot validly do more than predict the outcomes of observations, and that it should never purport to describe the reality that brings those outcomes about. This is known as *instrumentalism.* It denies that what I have been calling 'explanation' can exist at all. It is still very influential. In some fields (such as statistical analysis) the very word 'explanation' has come to mean prediction, so that a mathematical formula is said to 'explain' a set of experimental data. By 'reality' is meant merely the *observed data* that the formula is supposed to approximate. That leaves no term for assertions about reality itself, except perhaps 'useful fiction'.

Instrumentalism is one of many ways of denying *realism*, the commonsense, and true, doctrine that the physical world really exists, and is accessible to rational inquiry. Once one has denied this, the logical implication is that all claims about reality are equivalent to myths, none of them being better than the others in any objective sense. That is *relativism*, the doctrine that statements in a given field cannot be objectively true or false: at most they can be judged so relative to some cultural or other arbitrary standard.

Instrumentalism, even aside from the philosophical enormity of reducing science to a collection of statements about human experiences, does not make sense in its own terms. For there is no such thing as a purely predictive, explanationless theory. One cannot make even the simplest prediction without invoking quite a sophisticated explanatory framework. For example, those predictions about conjuring tricks apply specifically to conjuring tricks. That is explanatory information, and it tells me, among other things, not to 'extrapolate' the predictions to another type of situation, however successful they are at predicting conjuring tricks. So I know not to predict that saws in general are harmless to humans; and I continue to predict that if *I* were to place a ball under a cup, it really would go there and stay there.

The concept of a conjuring trick, and of the distinction between it and other situations, is familiar and unproblematic – so much so that it is easy to forget that it depends on substantive explanatory theories about all sorts of things such as how our senses work, how solid matter and light behave, and also subtle cultural details. Knowledge that is both familiar and uncontroversial is *background knowledge.* A predictive theory whose explanatory content consists only of background knowledge is a *rule of thumb*. Because we usually take background knowledge for granted, rules of thumb may seem to be explanationless predictions, but that is always an illusion.

There is always an explanation, whether we know it or not, for why a rule of thumb works. Denying that some regularity in nature has an explanation is effectively the same as believing in the supernatural – saying, 'That's not conjuring, it's actual magic.' Also, there is always an explanation when a rule of thumb *fails*, for rules of thumb are always parochial: they hold only in a narrow range of familiar circumstances. So, if an unfamiliar feature were introduced into a cups-and-balls trick, the rule of thumb I stated might easily make a false prediction. For instance, I could not tell from the rule of thumb whether it would be possible to perform the trick with lighted candles instead of balls. If I had an explanation of how the trick worked, I could tell.

Explanations are also essential for arriving at a rule of thumb in the first place: I could not have guessed those predictions about conjuring tricks without having a great deal of explanatory information in mind – even before any specific explanation of how the trick works. For instance, it is only in the light of explanations that I could have abstracted the concept of *cups* and *balls* from my experience of the trick, rather than, say, *red* and *blue*, even if it so happened that the cups were red and the balls blue in every instance of the trick that I had witnessed.

The essence of experimental testing is that there are at least two apparently viable theories known about the issue in question, making conflicting predictions that can be distinguished by the experiment. Just as conflicting predictions are the occasion for experiment and observation, so *conflicting ideas* in a broader sense are the occasion for all rational thought and inquiry. For example, if we are simply curious about something, it means that we believe that our existing ideas do not adequately capture or explain it. So, we have some *criterion* that our best existing explanation fails to meet. The criterion and the existing explanation are conflicting ideas. I shall call a situation in which we experience conflicting ideas a *problem*.

The example of a conjuring trick illustrates how observations provide problems for science – dependent, as always, on prior explanatory theories. For a conjuring trick is a trick only if it makes us think that *something happened* that *cannot happen*. Both halves of that proposition depend on our bringing quite a rich set of explanatory theories to the experience. That is why a trick that mystifies an adult may be uninteresting to a young child who has not yet learned to have the expectations on which the trick relies. Even those members of the audience who are incurious about how the trick works can detect that it *is* a trick only because of the explanatory theories that they brought with them into the auditorium. *Solving* a problem means creating an explanation that does not have the conflict.

Similarly, no one would have wondered what stars are if there had not been existing expectations – explanations – that unsupported things fall, and that lights need fuel, which runs out, and so on, which conflicted with interpretations (which are also explanations) of what was seen, such as that the stars shine constantly and do not fall. In this case it was those interpretations that were false: stars are indeed in free fall and do need fuel. But it took a great deal of conjecture, criticism and testing to discover how that can be.

A problem can also arise purely theoretically, without any observations. For instance, there is a problem when a theory makes a prediction that we did not expect. Expectations are theories too. Similarly, it is a problem when the way things *are* (according to our best explanation) is not the way they *should be* – that is, according to our current criterion of how they should be. This covers the whole range of ordinary meanings of the word 'problem', from unpleasant, as when the Apollo 13 mission reported, 'Houston, we've had a problem here,' to pleasant, as when Popper wrote:

> I think that there is only one way to science – or to philosophy, for that matter: to meet a problem, to see its beauty and fall in love with it; to get married to it and to live with it happily, till death do ye part – unless you should meet another and even more fascinating problem or unless, indeed, you should obtain a solution. But even if you do obtain a solution, you may then discover, to your delight, the existence of a whole family of enchanting, though perhaps difficult, problem children...
>
> *Realism and the Aim of Science* (1983)

Experimental testing involves many prior explanations in addition to the ones being tested, such as theories of how measuring instruments work. The refutation of a scientific theory has, from the point of view of someone who expected it to be true, the same logic as a conjuring trick – the only difference being that a conjurer does not normally have access to unknown laws of nature to make a trick work.

Since theories can contradict each other, but there are no contradictions in reality, every problem signals that our knowledge must be flawed or inadequate. Our misconception could be about the reality we are observing, or about how our perceptions are related to it, or both. For instance, a conjuring trick presents us with a problem only because we have misconceptions about what 'must' be happening – which implies that the knowledge that we used to interpret what we were seeing is defective. To an expert steeped in conjuring lore, it may be obvious what is happening – even if the expert did not observe the trick at all but merely heard a misleading account of it from a person who was fooled by it. This is another general fact about scientific explanation: if one has a misconception, observations that conflict with one's expectations may (or may not) spur one into making further conjectures, but no amount of observing will *correct* the misconception until after one has thought of a better idea; in contrast, if one has the right idea one can explain the phenomenon even if there are large errors in the data. Again, the very term 'data' ('givens') is misleading. Amending the 'data', or rejecting some as erroneous, is a frequent concomitant of scientific discovery, and the crucial 'data' cannot even be obtained until theory tells us what to look for and how and why.

A new conjuring trick is never totally unrelated to existing tricks. Like a new scientific theory, it is formed by creatively modifying, rearranging and combining the ideas from existing tricks. It requires pre-existing knowledge of how objects work and how audiences work, as well as how existing tricks work. So where did the earliest conjuring tricks come from? They must have been by the same observations. Yet what it asserts about reality is markedly different from – in many ways it is the opposite of – the original myth.

Every other detail of the story, apart from its bare prediction that winter happens once a year, is just as easily variable. So, although the myth was created to explain the seasons, it is only superficially adapted to that purpose. When its author was wondering what could possibly make a goddess do something once a year, he did not shout, 'Eureka! It must have been a marriage contract enforced by a magic seed.' He made that choice – and all his substantive choices as author – for cultural and artistic reasons, and not because of the attributes of winter at all. He may also have been trying to explain aspects of human nature metaphorically – but here I am concerned with the myth only in its capacity as an explanation *of seasons*, and in that respect even its author could not have denied that the role of all the details could be played equally well by countless other things.

The Persephone and Freyr myths assert radically incompatible things about what is happening in reality to cause seasons. Yet no one, I guess, has ever adopted either myth as a result of comparing it on its merits with the other, because there is no way of distinguishing between them. If we ignore all the parts of both myths whose role could be easily replaced, we are left with the same core explanation in both cases: *the gods did it*. Although Freyr is a very different god of spring from Persephone, and his battles very different events from her conjugal visits, none of those differing attributes has any function in the myths' respective accounts of why seasons happen. Hence none of them provides any reason for choosing one explanation over the other.

The reason those myths are so easily variable is that their details are barely connected to the details of the phenomena.

Nothing in the problem of why winter happens is addressed by postulating specifically a marriage contract or a magic seed, or the gods Persephone, Hades and Demeter – or Freyr. Whenever a wide range of variant theories can account equally well for the phenomenon they are trying to explain, there is no reason to prefer one of them over the others, so advocating a particular one in preference to the others is irrational.

That freedom to make drastic changes in those mythical explanations of seasons is the fundamental flaw in them. It is the reason that myth-making in general is not an effective way to understand the world. And that is so whether the myths are testable or not, for whenever it is easy to vary an explanation without changing its predictions, one could just as easily vary it to make different predictions if they were needed. For example, if the ancient Greeks *had* discovered that the seasons in the northern and southern hemispheres are out of phase, they would have had a choice of countless slight variants of the myth that would be consistent with that observation. One would be that when Demeter is sad she banishes warmth *from her vicinity*, and it has to go elsewhere – into the southern hemisphere. Similarly, slight variants of the Persephone explanation could account just as well for seasons that were marked by green rainbows, or seasons that happened once a week, or sporadically, or not at all. Likewise for the superstitious gambler or the end-of-the-world prophet: when their theory is refuted by experience, they do indeed switch to a new one; but, because their underlying explanations are bad, they can easily accommodate the new experience without changing the substance of the explanation. Without a good explanatory theory, they can simply reinterpret the omens, pick a new date, and make essentially the same prediction. In such cases, testing one's theory and abandoning it when it is refuted constitutes no progress towards understanding the world. If an explanation could easily explain anything in the given field, then it actually explains nothing.

In general, when theories are easily variable in the sense I have described, experimental testing is almost useless for correcting their errors. I call such theories *bad explanations.* Being proved wrong by experiment, and changing the theories to other bad explanations, does not get their holders one jot closer to the truth.

Because explanation plays this central role in science, and because testability is of little use in the case of bad explanations, I myself prefer to call myths, superstitions and similar theories *un*scientific even when they make testable predictions. But it does not matter what terminology you use, so long as it does not lead you to conclude that there is something worthwhile about the Persephone myth, or the prophet's apocalyptic theory or the gambler's delusion, just because it is testable. Nor is a person capable of making progress merely by virtue of being willing to drop a theory when it is refuted: one must also be seeking a better explanation of the relevant phenomena. That is the scientific frame of mind.

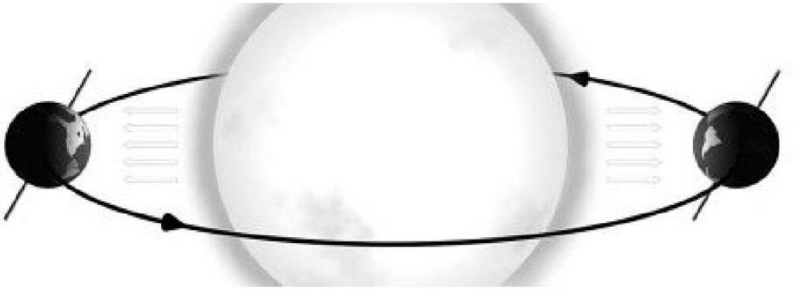
As the physicist Richard Feynman said, 'Science is what we have learned about how to keep from fooling ourselves.' By adopting easily variable explanations, the gambler and prophet are ensuring that they will be able to continue fooling themselves no matter what happens. Just as thoroughly as if they had adopted untestable theories, they are insulating themselves from facing evidence that they are mistaken about what is really there in the physical world.

The quest for good explanations is, I believe, the basic regulating principle not only of science, but of the Enlightenment generally. It is the feature that distinguishes those approaches to knowledge from all others, and it implies all those other conditions for scientific progress I have discussed: It trivially implies that prediction alone is insufficient. Somewhat less trivially, it leads to the rejection of authority, because if we adopt a theory on authority, that means that we would also have accepted a range of different theories on authority. And hence it also implies the need for a tradition of criticism. It also implies a methodological rule – a *criterion for reality* – namely that we should conclude that a particular thing is real if and only if it figures in our best explanation of something.

Although the pioneers of the Enlightenment and of the scientific revolution did not put it this way, seeking good explanations was (and remains) the spirit of the age. This is how they began to think. It is what they began to do, systematically for the first time. It is what made that momentous difference to the rate of progress of all kinds.

Long before the Enlightenment, there were individuals who sought good explanations. Indeed, my discussion here suggests that all progress then, as now, was due to such people. But in most ages they lacked contact with a tradition of criticism in which others could carry on their ideas, and so created little that left any trace for us to detect. We do know of sporadic traditions of good-explanation-seeking in narrowly defined fields, such as geometry, and even short-lived traditions of criticism – mini-enlightenments – which were tragically snuffed out, as I shall describe in Chapter 9. But the sea change in the values and patterns of thinking of a whole community of thinkers, which brought about a sustained and accelerating creation of knowledge, happened only once in history, with *the* Enlightenment and its scientific revolution. An entire political, moral, economic and intellectual culture – roughly what is now called 'the West' – grew around the values entailed by the quest for good explanations, such as tolerance of dissent, openness to change, distrust of dogmatism and authority, and the aspiration to progress both by individuals and for the culture as a whole. And the progress made by that multifaceted culture, in turn, promoted those values – though, as I shall explain in Chapter 15, they are nowhere close to being fully implemented.

Now consider the true explanation of seasons. It is that the Earth's axis of rotation is tilted relative to the plane of its orbit around the sun. Hence for half of each year the northern hemisphere is tilted towards the sun while the southern hemisphere is tilted away, and for the other half it is the other way around. Whenever the sun's rays are falling vertically in one hemisphere (thus providing more heat per unit area of the surface) they are falling obliquely in the other (thus providing less).



The true explanation of seasons (not to scale!)

That is a good explanation – hard to vary, because all its details play a functional role. For instance, we know – and can test independently of our experience of seasons – that surfaces tilted away from radiant heat are heated less than when they are facing it, and that a spinning sphere in space points in a constant direction. And we can explain why, in terms of theories of geometry, heat and mechanics. Also, the same tilt appears in our explanation of where the sun appears relative to the horizon at

Chapter 12.

When a formerly good explanation has been falsified by new observations, it is no longer a good explanation, because the problem has expanded to include those observations. Thus the standard scientific methodology of dropping theories when refuted by experiment is implied by the requirement for good explanations. The best explanations are the ones that are most constrained by existing knowledge – including other good explanations as well as other knowledge of the phenomena to be explained. That is why testable explanations that have passed stringent tests become extremely good explanations, which is in turn why the maxim of testability promotes the growth of knowledge in science.

Conjectures are the products of creative imagination. But the problem with imagination is that it can create fiction much more easily than truth. As I have suggested, historically, virtually all human attempts to explain experience in terms of a wider reality have indeed been fiction, in the form of myths, dogma and mistaken common sense – and the rule of testability is an insufficient check on such mistakes. But the quest for good explanations does the job: inventing falsehoods is easy, and therefore they are easy to vary once found; discovering good explanations is hard, but the harder they are to find, the harder they are to vary once found. The ideal that explanatory science strives for is nicely described by the quotation from Wheeler with which I began this chapter: 'Behind it all is surely an idea so simple, so beautiful, that when we grasp it – in a decade, a century, or a millennium – we will all say to each other, *how could it have been otherwise?* [*my italics*].' Now we shall see how this explanation-based conception of science answers the question that I asked above: how do we know so much about *unfamiliar* aspects of reality?

Put yourself in the place of an ancient astronomer thinking about the axis-tilt explanation of seasons. For the sake of simplicity, let us assume that you have also adopted the heliocentric theory. So you might be, say, Aristarchus of Samos, who gave the earliest known arguments for the heliocentric theory in the third century BCE.

Although you know that the Earth is a sphere, you possess no evidence about any location on Earth south of Ethiopia or north of the Shetland Islands. You do not know that there is an Atlantic or a Pacific ocean; to you, the known world consists of Europe, North Africa and parts of Asia, and the coastal waters nearby. Nevertheless, from the axis-tilt theory of seasons, you can make predictions about the weather in the literally unheard-of places beyond your known world. Some of these predictions are mundane and could be mistaken for induction: you predict that due east or west, however far you travel, you will experience seasons at about the same time of year (though the timings of sunrise and sunset will gradually shift with longitude). But you will also make some counter-intuitive predictions: if you travel only a little further north than the Shetlands, you will reach a frozen region where each day and each night last six months; if you travel further south than Ethiopia, you will first reach a place where there are no seasons, and then, still further south, you will reach a place where there are seasons, but they are perfectly out of phase with those everywhere in your known world. You have never travelled more than a few hundred kilometres from your home island in the Mediterranean. You have never experienced any seasons other than Mediterranean ones. You have never read, nor heard tell, of seasons that were out of phase with the ones you have experienced. But you know about them.

What if you'd rather not know? You may not like these predictions. Your friends and colleagues may ridicule them. *You may try to modify the explanation* so that it will not make them, without spoiling its agreement with observations and with other ideas for which you have no good alternatives. You will fail. That is what a good explanation will do for you: it makes it harder for you to fool yourself.

For instance, it may occur to you to modify your theory as follows: 'In the known world, the seasons happen at the times of year predicted by the axis-tilt theory; everywhere else on Earth, they *also* happen at those times of year.' This theory correctly predicts all evidence known to you. And it is just as testable as your real theory. But now, in order to deny what the axis-tilt theory predicts in the faraway places, you have had to deny what it says about reality, everywhere. The modified theory is no longer an explanation of seasons, just a (purported) rule of thumb. So denying that the original explanation describes the true cause of seasons in the places about which you have no evidence has forced you to deny that it describes the true cause even on your home island.

Suppose for the sake of argument that you thought of the axis-tilt theory yourself. It is your conjecture, your own original creation. Yet because it is a good explanation – hard to vary – it is not yours to modify. It has an autonomous meaning and an autonomous domain of applicability. You cannot confine its predictions to a region of your choosing. Whether you like it or not, it makes predictions about places both known to you and unknown to you, predictions that you have thought of and ones that you have not thought of. Tilted planets in similar orbits in other solar systems must have seasonal heating and cooling – planets in the most distant galaxies, and planets that we shall never see because they were destroyed aeons ago, and also planets that have yet to form. The theory reaches out, as it were, from its finite origins inside one brain that has been affected only by scraps of patchy evidence from a small part of one hemisphere of one planet – to infinity. This *reach* of explanations is another meaning of 'the beginning of infinity'. It is the ability of some of them to solve problems beyond those that they were created to solve.

The axis-tilt theory is an example: it was originally proposed to explain the changes in the sun's angle of elevation during each year. Combined with a little knowledge of heat and spinning bodies, it then explained seasons. And, without any further modification, it also explained why seasons are out of phase in the two hemispheres, and why tropical regions do not have them, and why the summer sun shines at midnight in polar regions – three phenomena of which its creators may well have been unaware.

The reach of an explanation is not a 'principle of induction'; it is not something that the creator of the explanation can use to obtain or justify it. It is not part of the creative process at all. We find out about it only after we have the explanation – sometimes long after. So it has nothing to do with 'extrapolation', or 'induction', or with 'deriving' a theory in any other alleged way. It is exactly the other way round: the reason that the explanation of seasons reaches far outside the experience of its creators is precisely that it *does not* have to be extrapolated. By its nature as an explanation, when its creators first thought of it, it already applied in our planet's other hemisphere, and throughout the solar system, and in other solar systems, and at other times.

Thus the reach of an explanation is neither an additional assumption nor a detachable one. It is determined by the content of the explanation itself. The better an explanation is, the more rigidly its reach is determined – because the harder it is to vary an explanation, the harder it is in particular to construct a variant with a different reach, whether larger or smaller, that is still an explanation. We expect the law of gravity to be the same on Mars as on Earth because only one viable explanation of gravity is known – Einstein's general theory of relativity – and that is a universal theory; but we do not expect the *map* of Mars to resemble the map of Earth, because our theories about how Earth looks, despite being excellent explanations, have no reach to the appearance of any other astronomical object. Always, it is explanatory theories that tell us which (usually few) aspects of one situation can be 'extrapolated' to others.

It also makes sense to speak of the reach of non-explanatory forms of knowledge – rules of thumb, and also knowledge that is implicit in the genes for biological adaptations. So, as I said, my rule of thumb about cups-and-balls tricks has reach to a certain class of tricks; but I could not know what that class is without the explanation for why the rule works.

Old ways of thought, which did not seek good explanations, permitted no process such as science for correcting errors and misconceptions. Improvements happened so rarely that most people never experienced one. Ideas were static for long periods. Being bad explanations, even the best of them typically had little reach and were therefore brittle and unreliable beyond, and often within, their traditional applications. When ideas did change, it was seldom for the better, and when it did happen to be for the better, that seldom increased their reach. The emergence of science, and more broadly what I am calling the Enlightenment, was the beginning of the end of such static, parochial systems of ideas. It initiated the present era in human history, unique for its sustained, rapid creation of knowledge with ever-increasing reach. Many have wondered how long this can continue. Is it inherently experienced.

Good/bad explanation An explanation that is hard/easy to vary while still accounting for what it purports to account for.

The Enlightenment (The beginning of) a way of pursuing knowledge with a tradition of criticism and seeking good explanations instead of reliance on authority.

Mini-enlightenment A short-lived tradition of criticism.

Rational Attempting to solve problems by seeking good explanations; actively pursuing error-correction by creating criticisms of both existing ideas and new proposals.

The West The political, moral, economic and intellectual culture that has been growing around the Enlightenment values of science, reason and freedom.

MEANINGS OF 'THE BEGINNING OF INFINITY' ENCOUNTERED IN THIS CHAPTER

– The fact that some explanations have reach.
– The universal reach of some explanations.
– The Enlightenment.
– A tradition of criticism.
– Conjecture: the origin of all knowledge.
– The discovery of how to make progress: science, the scientific revolution, seeking good explanations, and the political principles of the West.
– Fallibilism.

SUMMARY

Appearances are deceptive. Yet we have a great deal of knowledge about the vast and unfamiliar reality that causes them, and of the elegant, universal laws that govern that reality. This knowledge consists of explanations: assertions about what is out there beyond the appearances, and how it behaves. For most of the history of our species, we had almost no success in creating such knowledge. Where does it come from? Empiricism said that we derive it from sensory experience. This is false. The real source of our theories is conjecture, and the real source of our knowledge is conjecture alternating with criticism. We create theories by rearranging, combining, altering and adding to existing ideas with the intention of improving upon them. The role of experiment and observation is to choose between existing theories, not to be the source of new ones. We interpret experiences through explanatory theories, but true explanations are not obvious. Fallibilism entails not looking to authorities but instead acknowledging that we may always be mistaken, and trying to correct errors. We do so by seeking good explanations – explanations that are hard to vary in the sense that changing the details would ruin the explanation. This, not experimental testing, was the decisive factor in the scientific revolution, and also in the unique, rapid, sustained progress in other fields that have participated in the Enlightenment. That was a rebellion against authority which, unlike most such rebellions, tried not to seek authoritative justifications for theories, but instead set up a tradition of criticism. Some of the resulting ideas have enormous reach: they explain more than what they were originally designed to. The reach of an explanation is an intrinsic attribute of it, not an assumption that we make about it as empiricism and inductivism claim.

Now I'll say some more about appearance and reality, explanation and infinity.

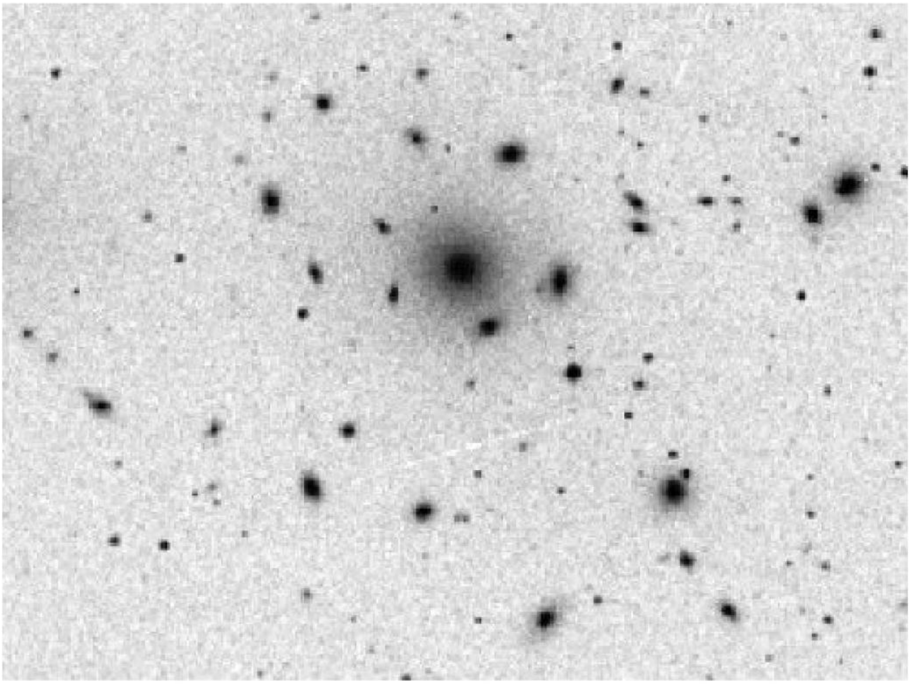
2

Closer to Reality

A galaxy is a mind-bogglingly huge thing. For that matter, a star is a mind-bogglingly huge thing. Our own planet is. A human brain is – in terms of both its internal complexity and the reach of human ideas. And there can be thousands of galaxies in a cluster, which can be millions of light years across. The phrase 'thousands of galaxies' trips lightly off the tongue, but it takes a while to make room in one's mind for the reality of it.

I was first stunned by the concept when I was a graduate student. Some fellow students were showing me what they were working on: observing clusters of galaxies – through *microscopes*. That is how astronomers used to use the Palomar Sky Survey, a collection of 1,874 photographic negatives of the sky, on glass plates, which showed the stars and galaxies as dark shapes on a white background.

They mounted one of the plates for me to look at. I focused the eyepiece of the microscope and saw something like this:



The Coma cluster of galaxies

Those fuzzy things are galaxies, and the sharply defined dots are stars in our own galaxy, thousands of times closer. The students' job was to catalogue the positions of the galaxies by lining them up in cross-hairs and pressing a button. I tried my hand at it – just for fun, since of course I was not qualified to make serious measurements. I soon found that it was not as easy as it had seemed. One reason is that it is not always obvious which are the galaxies and which are merely stars or other foreground objects. Some galaxies are easy to recognize: for instance, stars are never spiral, or noticeably elliptical. But some shapes are so faint that it is hard to tell whether they are sharp. Some galaxies appear small, faint and circular, and some are partly obscured by other objects. Nowadays such measurements are made by computers using learning anything. But, more profoundly, I expect that Edison was misinterpreting his own experience. A trial that fails is still fun. A repetitive experiment is not repetitive if one is thinking about the ideas that it is testing and the reality that it is investigating. That galaxy project was intended to discover whether 'dark matter' (see the next chapter) really exists – and it succeeded. If Edison, or those graduate students, or any scientific researcher engaged upon the 'perspiration' phase of discovery, had really been doing it mindlessly, they would be missing most of the fun – which is also what largely powers that 'one per cent inspiration'.

As I reached one particularly ambiguous image I asked my hosts, 'Is that a galaxy or a star?'

'Neither,' was the reply. 'That's just a defect in the photographic emulsion.'

The drastic mental gear change made me laugh. My grandiose speculations about the deep meaning of what I was seeing had turned out to be, in regard to this particular object, about nothing at all: suddenly there were no astronomers in that image, no rivers or earthquakes. They had disappeared in a puff of imagination. I had overestimated the mass of what I was looking at by some fifty powers of ten. What I had taken to be the largest object I had ever seen, and the most distant in space and time, was in reality just a speck barely visible without a microscope, within arm's reach. How easily, and how thoroughly, one can be misled.

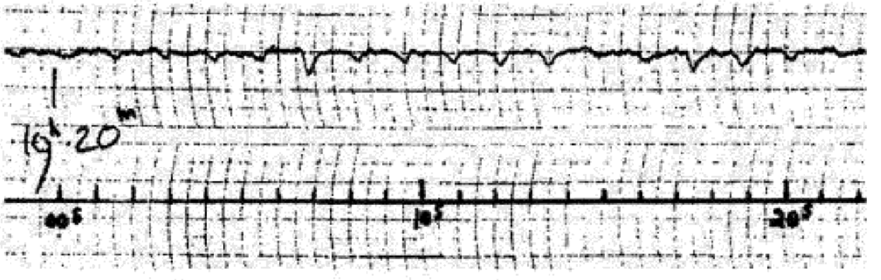
But wait. Was I ever looking at a galaxy? All the other blobs were in fact microscopic smudges of silver too. If I misclassified the *cause* of one of them, because it looked too like the others, why was that such a big error?

Because an error in experimental science *is* a mistake about the cause of something. Like an accurate observation, it is a matter of theory. Very little in nature is detectable by unaided human senses. Most of what happens is too fast or too slow, too big or too small, or too remote, or hidden behind opaque barriers, or operates on principles too different from anything that influenced our evolution. But in some cases we can arrange for such phenomena to become perceptible, via scientific instruments.

We experience such instruments as bringing us closer to the reality – just as I felt while looking at that galactic cluster. But in purely physical terms they only ever separate us further from it. I could have looked up at the night sky in the direction of that cluster, and there would have been nothing between it and my eye but a few grams of air – but I would have seen nothing at all. I could have interposed a telescope, and then I might have seen it. In the event, I was interposing a telescope, a camera, a photographic development laboratory, another camera (to make copies of the plates), a truck to bring the plates to my university, and a microscope. I could see the cluster far better with all that equipment in the way.

Astronomers nowadays never look up at the sky (except perhaps in their spare time), and hardly ever look through telescopes. Many telescopes do not even have eyepieces suitable for a human eye. Many do not even detect visible light. Instead, instruments detect invisible signals which are then digitized, recorded, combined with others, and processed and analysed by computers. As a result, images may be produced – perhaps in 'false colours' to indicate radio waves or other radiation, or to display still more indirectly inferred attributes such as temperature or composition. In many cases, no image of the distant object is ever produced, only lists of numbers, or graphs and diagrams, and only the outcome of those processes affects the astronomers' senses.

Every additional layer of physical separation requires further levels of theory to relate the resulting perceptions to reality. When the astronomer Jocelyn Bell discovered pulsars (extremely dense stars that emit regular bursts of radio waves), this is what she was looking at:



Radio-telescope output from the first known pulsar

Only through a sophisticated chain of theoretical interpretation could she ‘see’, by looking at that shaky line of ink on paper, a powerful, pulsating object in deep space, and recognize that it was of a hitherto unknown type.

The better we come to understand phenomena remote from our everyday experience, the longer those chains of interpretation become, and every additional link necessitates more theory. A single unexpected or misunderstood phenomenon anywhere in the chain can, and often does, render the resulting sensory experience arbitrarily misleading. Yet, over time, the conclusions that science has drawn have become ever truer to reality. Its quest for good explanations corrects the errors, allows for the biases and misleading perspectives, and fills in the gaps. This is what we can achieve when, as Feynman said, we keep learning more about how not to fool ourselves.

Telescopes contain automatic tracking mechanisms that continuously realign them so as to compensate for the effect of the Earth’s motion; in some, computers continuously change the shape of the mirror so as to compensate for the shimmering of the Earth's atmosphere. And so, observed through such a telescope, stars do not appear to twinkle or to move across the sky as they did to generations of observers in the past. Those things are only appearance – parochial error. They have nothing to do with the reality of stars. The primary function of the telescope's optics is to reduce the illusion that the stars are few, faint, twinkling and moving. The same is true of every feature of the telescope, and of all other scientific instruments: each layer of indirectness, through its associated theory, corrects errors, illusions, misleading perspectives and gaps. Perhaps it is the mistaken empiricist ideal of 'pure', theory-free observation that makes it seem odd that truly accurate observation is always so hugely indirect. But the fact is that progress requires the application of ever more knowledge *in advance* of our observations.

So I was indeed looking at galaxies. Observing a galaxy via specks of silver is no different in that regard from observing a garden via images on a retina. In all cases, to say that we have genuinely observed any given thing is to say that we have accurately attributed our evidence (ultimately always evidence inside our own brains) to that thing. Scientific truth consists of such correspondence between theories and physical reality.

Scientists operating giant particle accelerators likewise look at pixels and ink, numbers and graphs, and thereby observe the microscopic reality of subatomic particles like nuclei and quarks. Others operate electron microscopes and fire the beam at cells that are as dead as dodos, having been stained, quick-frozen by liquid nitrogen, and mounted in a vacuum – but they thereby learn what *living* cells are like. It is a marvellous fact that objects can exist which, when we observe them, accurately take on the appearance and other attributes of other objects that are elsewhere and very differently constituted. Our sensory systems are such objects too, for it is only they that are directly affecting our brains when we perceive anything.

Such instruments are rare and fragile configurations of matter. Press one wrong button on the telescope's control panel, or code one wrong instruction into its computer, and the whole immensely complex artefact may well revert to revealing nothing other than itself. The same would be true if, instead of making that scientific instrument, you were to assemble those raw materials into almost any other configuration: stare at them, and you would see nothing other than them.

Explanatory theories tell us how to build and operate instruments in exactly the right way to work this miracle. Like conjuring tricks in reverse, such instruments fool our senses into seeing what is really there. Our minds, through the methodological criterion that I mentioned in Chapter 1, conclude that a particular thing is real if and only if it figures in our best explanation of something. Physically, all that has happened is that human beings, on Earth, have dug up raw materials such as iron ore and sand, and have rearranged them – still on Earth – into complex objects such as radio telescopes, computers and display screens, and now, instead of looking at the sky, they look at those objects. They are focusing their *eyes* on human artefacts that are close enough to touch. But their *minds* are focused on alien entities and processes, light years away.

Sometimes they are still looking at glowing dots just as their ancestors did – but on computer monitors instead of the sky. Sometimes they are looking at numbers or graphs. But in all cases they are inspecting local phenomena: pixels on a screen, ink on paper, and so on. These things are physically very unlike stars: they are much smaller; they are not dominated by nuclear forces

3

The Spark

Most ancient accounts of the reality beyond our everyday experience were not only false, they had a radically different character from modern ones: they were *anthropocentric*. That is to say, they centred on human beings, and more broadly on *people* – entities with intentions and human-like thoughts – which included powerful, supernatural people such as spirits and gods. So, winter might be attributed to someone's sadness, harvests to someone's generosity, natural disasters to someone's anger, and so on. Such explanations often involved cosmically significant beings caring what humans did, or having intentions about them. This conferred cosmic significance on humans too. Then the geocentric theory placed humans at the physical hub of the universe as well. Those two kinds of anthropocentrism – explanatory and geometrical – made each other more plausible, and, as a result, pre-Enlightenment thinking was more anthropocentric than we can readily imagine nowadays.

A notable exception was the science of geometry itself, especially the system developed by the ancient Greek mathematician Euclid. Its elegant axioms and modes of reasoning about impersonal entities such as points and lines would later be an inspiration to many of the pioneers of the Enlightenment. But until then it had little impact on prevailing world views. For example, most astronomers were also astrologers: despite using sophisticated geometry in their work, they believed that the stars foretold political and personal events on Earth.

Before anything was known about how the world works, trying to explain physical phenomena in terms of purposeful, human-like thought and action may have been a reasonable approach. After all, that is how we explain much of our everyday experience even today: if a jewel is mysteriously missing from a locked safe, we seek human-level explanations such as error or theft (or, under some circumstances, conjuring), not new laws of physics. But that anthropocentric approach has never yielded any good explanations beyond the realm of human affairs. In regard to the physical world at large, it was colossally misconceived. We now know that the patterns of stars and planets in our night sky have no significance for human affairs. We know that we are not at the centre of the universe – it does not even have a geometrical centre. And we know that, although some of the titanic astrophysical phenomena that I have described played a significant role in our past, we have never been significant to them. We call a phenomenon *significant* (or *fundamental*) if parochial theories are inadequate to explain it, or if it appears in the explanation of many other phenomena; so it may seem that human beings and their wishes and actions are extremely insignificant in the universe at large.

Anthropocentric misconceptions have also been overturned in every other fundamental area of science: our knowledge of physics is now expressed entirely in terms of entities that are as impersonal as Euclid's points and lines, such as elementary particles, forces and spacetime – a four-dimensional continuum with three dimensions of space and one of time. Their effects on each other are explained not in terms of feelings and intentions, but through mathematical equations expressing laws of nature. In biology, it was once thought that living things must have been designed by a supernatural person, and that they must contain some special ingredient, a 'vital principle', to make them behave with apparent purposefulness. But biological science discovered new modes of explanation through such impersonal things as chemical reactions, genes and evolution. So we now know that living things, including humans, all consist of the same ingredients as rocks and stars, and obey the same laws, and that they were not designed by anyone. Modern science, far from explaining physical phenomena in terms of the thoughts and intentions of unseen people, considers our own thoughts and intentions to be aggregates of unseen (though not unseeable) microscopic physical processes in our brains.

So fruitful has this abandonment of anthropocentric theories been, and so important in the broader history of ideas, that *anti*-anthropocentrism has increasingly been elevated to the status of a universal principle, sometimes called the 'Principle of Mediocrity': *there is nothing significant about humans (in the cosmic scheme of things)*. As the physicist Stephen Hawking put it, humans are 'just a chemical scum on the surface of a typical planet that's in orbit round a typical star on the outskirts of a typical galaxy'. The proviso 'in the cosmic scheme of things' is necessary because the chemical scum evidently does have a special significance according to values that it applies to itself, such as moral values. But the Principle says that all such values are themselves anthropocentric: they explain only the behaviour of the scum, which is itself insignificant.

It is easy to mistake quirks of one's own, familiar environment or perspective (such as the rotation of the night sky) for objective features of what one is observing, or to mistake rules of thumb (such as the prediction of daily sunrises) for universal laws. I shall refer to that sort of error as *parochialism*.

Anthropocentric errors are examples of parochialism, but not all parochialism is anthropocentric. For instance, the prediction that the seasons are in phase all over the world is a parochial error but not an anthropocentric one: it does not involve explaining seasons in terms of people.

Another influential idea about the human condition is sometimes given the dramatic name *Spaceship Earth*. Imagine a 'generation ship' – a spaceship on a journey so long that many generations of passengers live out their lives in transit. This has been proposed as a means of colonizing other star systems. In the Spaceship Earth idea, that generation ship is a metaphor for the *biosphere* – the system of all living things on Earth and the regions they inhabit. Its passengers represent all humans on Earth. Outside the spaceship, the universe is implacably hostile, but the interior is a vastly complex life-support system, capable of providing everything that the passengers need to thrive. Like the spaceship, the biosphere recycles all waste and, using its capacious nuclear power plant (the sun), it is completely self-sufficient.

Just as the spaceship's life-support system is designed to sustain its passengers, so the biosphere has the 'appearance of design': it seems highly adapted to sustaining us (claims the metaphor) because *we* were adapted to *it* by evolution. But its capacity is finite: if we overload it, either by our sheer numbers or by adopting lifestyles too different from those that we evolved to live (the ones that it is 'designed' to support), it will break down. And, like the passengers on that spaceship, we get no second chances: if our lifestyle becomes too careless or profligate and we ruin our life-support system, we have nowhere else to go.

The Spaceship Earth metaphor and the Principle of Mediocrity have both gained wide acceptance among scientifically minded people – to the extent of becoming truisms. This is despite the fact that, on the face of it, they argue in somewhat opposite directions: the Principle of Mediocrity stresses how *typical* the Earth and its chemical scum are (in the sense of being unremarkable), while Spaceship Earth stresses how *untypical* they are (in the sense of being uniquely suited to each other). But when the two ideas are interpreted in broad, philosophical ways, as they usually are, they can easily converge. Both see themselves as correcting much the same parochial misconceptions, namely that our experience of life on Earth is representative of the universe, and that the Earth is vast, fixed and permanent. They both stress instead that it is tiny and ephemeral. Both oppose arrogance: the Principle of Mediocrity opposes the pre-Enlightenment arrogance of believing ourselves significant in the world; the Spaceship Earth metaphor opposes the Enlightenment arrogance of aspiring to control the world. Both have a moral element: we *should not* consider ourselves significant, they assert; we *should not* expect the world to submit indefinitely to our depredations.

Thus the two ideas generate a rich conceptual framework that can inform an entire world view. Yet, as I shall explain, they are both false, even in the straightforward factual sense. And in the broader sense they are so misleading that, if you were seeking maxims worth being carved in stone and recited each morning before breakfast, you could do a lot worse than to use their *negations*. That is to say, the truth is that

People *are* significant in the cosmic scheme of things; and
The Earth's biosphere is *incapable* of supporting human life.

Consider Hawking's remark again. It is true that we are on a (somewhat) typical planet of a typical star in a typical galaxy. But we are far from typical of the matter in the universe. For one thing, about 80 per cent of that matter is thought to be invisible 'dark matter', which can neither emit nor absorb light. We known substance except helium. (Helium is believed to remain liquid right down to absolute zero, unless highly pressurized.)

And it is empty: the density of atoms out there is below one per cubic metre. That is a million times sparser than atoms in the space between the stars, and those atoms are themselves sparser than in the best vacuum that human technology has yet achieved. Almost all the atoms in intergalactic space are hydrogen or helium, so there is no chemistry. No life could have evolved there, nor any intelligence. Nothing changes there. Nothing happens. The same is true of the next cube and the next, and if you were to examine a million consecutive cubes in any direction the story would be the same.

Cold, dark and empty. That unimaginably desolate environment is typical of the universe – and is another measure of how *un*typical the Earth and its chemical scum are, in a straightforward physical sense. The issue of the cosmic significance of this type of scum will shortly take us back out into intergalactic space. But let me first return to Earth, and consider the Spaceship Earth metaphor, in *its* straightforward physical version.

This much is true: if, tomorrow, physical conditions on the Earth's surface were to change even slightly by astrophysical standards, then no humans could live here unprotected, just as they could not survive on a spaceship whose life-support system had broken down. Yet I am writing this in Oxford, England, where winter nights are likewise often cold enough to kill any human unprotected by clothing and other technology. So, while intergalactic space would kill me in a matter of seconds, Oxfordshire in its primeval state might do it in a matter of hours – which can be considered 'life support' only in the most contrived sense. There *is* a life-support system in Oxfordshire today, but it was not provided by the biosphere. It has been built by humans. It consists of clothes, houses, farms, hospitals, an electrical grid, a sewage system and so on. Nearly the whole of the Earth's biosphere in its primeval state was likewise incapable of keeping an unprotected human alive for long. It would be much more accurate to call it a death trap for humans rather than a life-support system. Even the Great Rift Valley in eastern Africa, where our species evolved, was barely more hospitable than primeval Oxfordshire. Unlike the life-support system in that imagined spaceship, the Great Rift Valley lacked a safe water supply, and medical equipment, and comfortable living quarters, and was infested with predators, parasites and disease organisms. It frequently injured, poisoned, drenched, starved and sickened its 'passengers', and most of them died as a result.

It was similarly harsh to all the other organisms that lived there: few individuals live comfortably or die of old age in the supposedly beneficent biosphere. That is no accident: most populations, of most species, are living close to the edge of disaster and death. It has to be that way, because as soon as some small group, somewhere, begins to have a slightly easier life than that, for any reason – for instance, an increased food supply, or the extinction of a competitor or predator – then its numbers increase. As a result, its other resources are depleted by the increased usage; so an increasing proportion of the population now has to colonize more marginal habitats and make do with inferior resources, and so on. This process continues until the disadvantages caused by the increased population have exactly balanced the advantage conferred by the beneficial change. That is to say, the new birth rate is again just barely keeping pace with the rampant disabling and killing of individuals by starvation, exhaustion, predation, overcrowding and all those other natural processes.

That is the situation to which evolution adapts organisms. And that, therefore, is the lifestyle in which the Earth's biosphere 'seems adapted' to sustaining them. The biosphere only ever achieves stability – and only temporarily at that – by continually neglecting, harming, disabling and killing individuals. Hence the metaphor of a spaceship or a life-support system, is quite perverse: when humans design a life-support system, they design it to provide the maximum possible comfort, safety and longevity for its users within the available resources; the biosphere has no such priorities.

Nor is the biosphere a great preserver of *species*. In addition to being notoriously cruel to individuals, evolution involves continual extinctions of entire species. The average rate of extinction since the beginning of life on Earth has been about ten species per year (the number is known only very approximately), becoming much higher during the relatively brief periods that palaeontologists call 'mass extinction events'. The rate at which species have come into existence has on balance only slightly exceeded the extinction rate, and the net effect is that the overwhelming majority of species that have ever existed on Earth (perhaps 99.9 per cent of them) are now extinct. Genetic evidence suggests that our own species narrowly escaped extinction on at least one occasion. Several species closely related to ours did become extinct. Significantly, the 'life-support system' itself wiped them out – by means such as natural disasters, evolutionary changes in other species, and climate change. Those cousins of ours had not invited extinction by changing their lifestyles or overloading the biosphere: on the contrary, it wiped them out because they *were* living the lifestyles that they had evolved to live, and in which, according to the Spaceship Earth metaphor, the biosphere had been 'supporting' them.

Yet that still overstates the degree to which the biosphere is hospitable to humans in particular. The first people to live at the latitude of Oxford (who were actually from a species related to us, possibly the Neanderthals) could do so only because they brought knowledge with them, about such things as tools, weapons, fire and clothing. That knowledge was transmitted from generation to generation not genetically but culturally. Our pre-human ancestors in the Great Rift Valley used such knowledge too, and our own species must have come into existence already dependent on it for survival. As evidence of that, note that I would soon die if I tried to live in the Great Rift Valley in its primeval state: I do not have the requisite knowledge. Since then, there have been human populations who, for instance, knew how to survive in the Amazon jungle but not in the Arctic, and populations for whom it was the other way round. Therefore that knowledge was not part of their genetic inheritance. It was created by human thought, and preserved and transmitted in human culture.

Today, almost the entire capacity of the Earth's 'life-support system for humans' has been provided not *for* us but *by* us, using our ability to create new knowledge. There are people in the Great Rift Valley today who live far more comfortably than early humans did, and in far greater numbers, through knowledge of things like tools, farming and hygiene. The Earth did provide the raw materials for our survival – just as the sun has provided the energy, and supernovae provided the elements, and so on. But a heap of raw materials is not the same thing as a life-support system. It takes knowledge to convert the one into the other, and biological evolution never provided us with enough knowledge to survive, let alone to thrive. In this respect we differ from almost all other species. They do have all the knowledge that they need, genetically encoded in their brains. And that knowledge was indeed provided for them by evolution – and so, in the relevant sense, 'by the biosphere'. So *their* home environments do have the appearance of having been designed as life-support systems for them, albeit only in the desperately limited sense that I have described. But the biosphere no more provides humans with a life-support system than it provides us with radio telescopes.

So the biosphere is incapable of supporting human life. From the outset, it was only human knowledge that made the planet even marginally habitable by humans, and the enormously increased capacity of our life-support system since then (in terms both of numbers and of security and quality of life) has been entirely due to the creation of human knowledge. To the extent that we are on a 'spaceship', we have never been merely its passengers, nor (as is often said) its stewards, nor even its maintenance crew: we are its designers and builders. Before the designs created by humans, it was not a vehicle, but only a heap of dangerous raw materials.

The 'passengers' metaphor is a misconception in another sense too. It implies that there was a time when humans lived unproblematically: when they were provided for, like passengers, without themselves having to solve a stream of problems in order to survive and to thrive. But in fact, even with the benefit of their cultural knowledge, our ancestors continually faced desperate problems, such as where the next meal was coming from, and typically they barely solved these problems or they died. There are very few fossils of old people.

The moral component of the Spaceship Earth metaphor is therefore somewhat paradoxical. It casts humans as ungrateful for gifts which, in reality, they never received. And it casts all other species in morally positive roles in the spaceship's life-support system, with humans as the only negative actors. But humans are agrees with an earlier evolutionary biologist, John Haldane, who expected that 'the universe is not only queerer than we suppose, but queerer than we *can* suppose.'

That is a startling – and paradoxical – consequence of the Principle of Mediocrity: it says that all human abilities, including the distinctive ones such as the ability to create new explanations, are necessarily parochial. That implies, in particular, that progress in science cannot exceed a certain limit defined by the biology of the human brain. And we must expect to reach that limit sooner rather than later. Beyond it, the world stops making sense (or seems to). The answer to the question that I asked at the end of Chapter 2 – whether the scientific revolution and the broader Enlightenment could be a beginning of infinity – would then be a resounding no. Science, for all its successes and aspirations, would turn out to be inherently parochial – and, ironically, anthropocentric.

So here the Principle of Mediocrity and Spaceship Earth converge. They share a conception of a tiny, human-friendly bubble embedded in the alien and uncooperative universe. The Spaceship Earth metaphor sees it as a physical bubble, the biosphere. For the Principle of Mediocrity, the bubble is primarily conceptual, marking the limits of the human capacity to understand the world. Those two bubbles are related, as we shall see. In both views, anthropocentrism is true in the interior of the bubble: there the world is unproblematic, uniquely compliant with human wishes and human understanding. Outside it there are only insoluble problems.

Dawkins would prefer it to be otherwise. As he wrote:

> I believe that an orderly universe, one indifferent to human preoccupations, in which everything has an explanation

> even if we still have a long way to go before we find it, is a more beautiful, more wonderful place than a universe tricked out with capricious ad hoc magic.
>
> *Unweaving the Rainbow* (1998)

An 'orderly' (explicable) universe is indeed more beautiful (see Chapter 14) – though the assumption that to be orderly it has to be 'indifferent to human preoccupations' is a misconception associated with the Principle of Mediocrity.

Any assumption that the world is *in*explicable can lead only to extremely bad explanations. For an inexplicable world is indistinguishable from one 'tricked out with capricious ad hoc magic': by definition, no hypothesis about the world outside the bubble of explicability can be a better explanation than that Zeus rules there – or practically any myth or fantasy one likes.

Moreover, since the outside of the bubble affects our explanations of the inside (or else we may as well do without it), the inside is not really explicable either. It seems so only if we carefully refrain from asking certain questions. This bears an uncanny resemblance to the intellectual landscape before the Enlightenment, with its distinction between Earth and heaven. It is a paradox inherent in the Principle of Mediocrity: contrary to its motivation, here it is forcing us back to an archaic, anthropocentric, pre-scientific conception of the world.

At root, the Principle of Mediocrity and the Spaceship Earth metaphor overlap in a claim about *reach*: they both claim that the reach of the distinctively human way of being – that is to say, the way of problem-solving, knowledge-creating and adapting the world around us – is bounded. And they argue that its bounds cannot be very far beyond what it has already reached. Trying to go beyond that range must lead to failure and catastrophe respectively.

Both ideas also rely on essentially the same argument, namely that if there were no such limit, there would be no explanation for the continued effectiveness of the adaptations of the human brain beyond the conditions under which they evolved. Why should one adaptation out of the trillions that have ever existed on Earth have unlimited reach, when all others reach only inside the tiny, insignificant, untypical biosphere? Fair enough: all reach has an explanation. But what if there *is* an explanation, and what if it has nothing to do with evolution or the biosphere?

Imagine that a flock of birds from a species that evolved on one island happens to fly to another. Their wings and eyes still work. That is an example of the reach of those adaptations. It has an explanation, the essence of which is that wings and eyes exploit universal laws of physics (of aerodynamics and optics respectively). They exploit those laws only imperfectly; but the atmospheric and lighting conditions on the two islands are sufficiently similar, by the criteria defined by those laws, for the same adaptations to work on both.

Thus the birds may well be able to fly to an island many kilometres away horizontally, but if they were transported only a few kilometres upwards their wings would stop working because the density of the air would be too low. Their implicit knowledge about how to fly fails at high altitude. A little further up, their eyes and other organs would stop working. The design of these too does not have that much reach: all vertebrate eyes are filled with liquid water, but water freezes at stratospheric temperatures and boils in the vacuum of space. Less dramatically, the birds might also die if they merely had no good night vision and they reached an island where the only suitable prey organisms were nocturnal. For the same reason, biological adaptations also have limited reach in regard to changes in their *home* environment – which can and do cause extinctions.

If those birds' adaptations do have enough reach to make the species viable on the new island, they will set up a colony there. In subsequent generations, mutants slightly better adapted to the new island will end up having slightly more offspring on average, so evolution will adapt the population more accurately to contain the knowledge needed to make a living there. The ancestor species of humans colonized new habitats and embarked on new lifestyles in exactly that way. But by the time our species had evolved, our fully human ancestors were achieving much the same thing thousands of times faster, by evolving their cultural knowledge instead. Because they did not yet know how to do science, their knowledge was only a little less parochial than biological knowledge. It consisted of rules of thumb. And so progress, though rapid compared to biological evolution, was sluggish compared to what the Enlightenment has accustomed us to.

Since the Enlightenment, technological progress has depended specifically on the creation of explanatory knowledge. People had dreamed for millennia of flying to the moon, but it was only with the advent of Newton's theories about the behaviour of invisible entities such as forces and momentum that they began to understand what was needed in order to go there.

This increasingly intimate connection between *explaining* the world and *controlling* it is no accident, but is part of the deep structure of the world. Consider the set of all conceivable transformations of physical objects. Some of those (like faster-than-light communication) never happen because they are forbidden by laws of nature; some (like the formation of stars out of primordial hydrogen) happen spontaneously; and some (such as converting air and water into trees, or converting raw materials into a radio telescope) are possible, but happen only when the requisite knowledge is present – for instance, embodied in genes or brains. But those are the only possibilities. That is to say, every putative physical transformation, to be performed in a given time with given resources or under any other conditions, is either

– impossible because it is forbidden by the laws of nature; or
– achievable, given the right knowledge.

That momentous dichotomy exists because if there were transformations that technology could never achieve regardless of what knowledge was brought to bear, then this fact would itself be a testable regularity in nature. But all regularities in nature have explanations, so the explanation of that regularity would itself be a law of nature, or a consequence of one. And so, again, everything that is not forbidden by laws of nature is achievable, given the right knowledge.

This fundamental connection between explanatory knowledge and technology is why the Haldane–Dawkins queerer-than-we-can-suppose argument is mistaken – why the reach of human adaptations does have a different character from that of all the other adaptations in the biosphere. The ability to create and use explanatory knowledge gives *people* a power to transform nature which is ultimately not limited by parochial factors, as all other adaptations are, but only by universal laws. This is the cosmic significance of explanatory knowledge – and hence of people, whom I shall henceforward define as entities that can create explanatory knowledge.

For every other species on Earth, we can determine its reach simply by making a list of all the resources and environmental conditions on which its adaptations depend. In principle one could

Eventually the moon colonists will take air for granted, just as the people now living in Oxfordshire take for granted that water will flow if they turn on a tap. If either of those populations lacked the right knowledge, their environment would soon kill them.

We are accustomed to thinking of the Earth as hospitable and the moon as a bleak, faraway deathtrap. But that is how our ancestors would have regarded Oxfordshire, and, ironically, it is how I, today, would regard the primeval Great Rift Valley. In the unique case of humans, the difference between a hospitable environment and a deathtrap depends on what knowledge they have created. Once enough knowledge has been embodied in the lunar colony, the colonists can devote their thoughts and energies to creating even more knowledge, and soon it will cease to be a colony and become simply home. No one will think of the moon as a fringe habitat, distinguished from our 'natural' environment on Earth, any more than we now think of Oxfordshire as being fundamentally different from the Great Rift Valley as a place to live.

Using knowledge to cause automated physical transformations is, in itself, not unique to humans. It is the basic method by which all organisms keep themselves alive: every cell is a chemical factory. The difference between humans and other species is in what kind of knowledge they can use (explanatory instead of rule-of-thumb) and in how they create it (conjecture and criticism of ideas, rather than the variation and selection of genes). It is precisely those two differences that explain why every other organism can function only in a certain range of environments that are hospitable to it, while humans transform *in*hospitable environments like the biosphere into support systems for themselves. And, while every other organism is a factory for converting resources of a fixed type into more such organisms, human bodies (including their brains) are factories for transforming *anything into anything* that the laws of nature allow. They are 'universal constructors'.

This universality in the human condition is part of a broader phenomenon that I shall discuss in Chapter 6. We do not share it with any other species currently on Earth. But, since it is a consequence of the ability to create explanations, we do necessarily share it with any other people that might exist in the universe. The opportunities provided by the laws of nature for transforming resources are universal, and all entities with universal reach necessarily have the same reach.

A few species other than humans are known to be capable of having cultural knowledge. For example, some apes can discover new methods of cracking nuts, and pass that knowledge on to other apes. As I shall discuss in Chapter 16, the existence of such knowledge is suggestive of how ape-like species evolved into people. But it is irrelevant to the arguments of this chapter, because no such organism is capable of creating or using explanatory knowledge. Hence the cultural knowledge of such organisms is of essentially the same type as genetic knowledge, and does indeed have only a small and inherently limited reach. They are not universal constructors, but highly specialized ones. For them, the Haldane–Dawkins argument is valid: the world is stranger than they can conceive.

In some environments in the universe, the most efficient way for humans to thrive might be to alter their own genes. Indeed, we are already doing that in our present environment, to eliminate diseases that have in the past blighted many lives. Some people object to this on the grounds (in effect) that a genetically altered human is no longer human. This is an anthropomorphic mistake. The only uniquely significant thing about humans (whether in the cosmic scheme of things or according to any rational human criterion) is our ability to create new explanations, and we have that in common with all people. You do not become less of a person if you lose a limb in an accident; it is only if you lose your brain that you do. Changing our genes in order to improve our lives and to facilitate further improvements is no different in this regard from augmenting our skin with clothes or our eyes with telescopes.

One might wonder whether the reach of people in general might be greater than the reach of humans. What if, for instance, the reach of technology is indeed unlimited, but only to creatures with two opposable thumbs on each hand; or if the reach of scientific knowledge is unlimited, but only to beings whose brains are twice the size of ours? But our faculty of being universal constructors makes these issues as irrelevant as that of access to vitamins. If progress at some point were to depend on having two thumbs per hand, then the outcome would depend not on the knowledge we inherit in our genes, but on whether we could discover how to build robots, or gloves, with two thumbs per hand, or alter ourselves to have a second thumb. If it depends on having more memory capacity, or speed, than a human brain, then the outcome would depend on whether we could build computers to do the job. Again, such things are already commonplace in technology.

The astrophysicist Martin Rees has speculated that somewhere in the universe 'there could be life and intelligence out there in forms we can't conceive. Just as a chimpanzee can't understand quantum theory, it could be there are aspects of reality that are beyond the capacity of our brains.' But that cannot be so. For if the 'capacity' in question is mere computational speed and amount of memory, then we can understand the aspects in question with the help of computers – just as we have understood the world for centuries with the help of pencil and paper. As Einstein remarked, 'My pencil and I are more clever than I.' In terms of computational repertoire, our computers – and brains – are already universal (see Chapter 6). But if the claim is that we may be *qualitatively* unable to understand what some other forms of intelligence can – if our disability cannot be remedied by mere automation – then this is just another claim that the world is not explicable. Indeed, it is tantamount to an appeal to the supernatural, with all the arbitrariness that is inherent in such appeals, for if we wanted to incorporate into our world view an imaginary realm explicable only to superhumans, we need never have bothered to abandon the myths of Persephone and her fellow deities.

So human reach is essentially the same as the reach of explanatory knowledge itself. An environment is within human reach if it is possible to create an open-ended stream of explanatory knowledge there. That means that if knowledge of a suitable kind were instantiated in such an environment in suitable physical objects, it would cause itself to survive and would then continue to increase indefinitely. Can there really be such an environment? This is essentially the question that I asked at the end of the last chapter – *can this creativity continue indefinitely?* – and it is the question to which the Spaceship Earth metaphor assumes a negative answer.

The issue comes down to this: if such an environment can exist, what are the minimal physical features that it must have? Access to *matter* is one. For example, the trick of extracting oxygen from moon rocks depends on having compounds of oxygen available. With more advanced technology, one could manufacture oxygen by transmutation; but, no matter how advanced one's technology is, one still needs raw materials of some sort. And, although mass can be recycled, creating an open-ended stream of knowledge depends on having an ongoing supply of it, both to make up for inevitable inefficiencies and to make the additional memory capacity to store new knowledge as it is created.

Also, many of the necessary transformations require *energy*: something must power conjectures and scientific experiments and all those manufacturing processes; and, again, the laws of physics forbid the creation of energy from nothing. So access to an energy supply is also a necessity. To some extent, energy and mass can be transformed into each other. For instance, transmuting hydrogen into any other element releases energy through nuclear fusion. Energy can also be converted into mass by various subatomic processes (but I cannot imagine naturally occurring circumstances in which those would be the best way of obtaining matter).

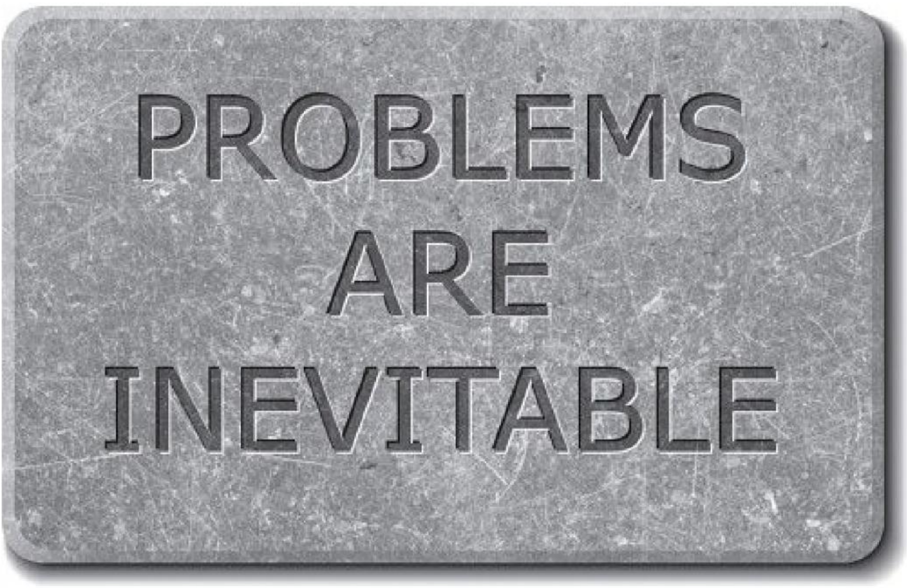
In addition to matter and energy, there is one other essential requirement, namely *evidence*: the information needed to test scientific theories. The Earth's surface is rich in evidence. We happened to get round to testing Newton's laws in the seventeenth century, and Einstein's in the twentieth, but the evidence with which we did that – light from the sky – had been deluging the surface of the Earth for billions of years before that, and will continue to do so for billions more. Even today we have barely begun to examine that evidence: on any clear night, the chances are that your roof will be struck by evidence falling from the sky which, if you only knew what to look for and how, would win you a Nobel prize. In chemistry, every stable element that exists anywhere is also present on or just below the Earth's surface. In biology, copious evidence of the nature of life is ubiquitous in the biosphere – and within arm's reach, in our own DNA. As far as we know, all the fundamental constants of nature can be measured here, and every fundamental law can be tested here. Everything

In fact people will always want still more than that: they will want to make progress. For, in addition to threats, there will always be problems in the benign sense of the word: errors, gaps, inconsistencies and inadequacies in our knowledge that we wish to solve – including, not least, moral knowledge: knowledge about *what to want*, what to strive for. The human mind seeks explanations; and now that we know how to find them, we are not going to stop voluntarily. Here is another misconception in the Garden of Eden myth: that the supposed unproblematic state would be a *good* state to be in. Some theologians have denied this, and I agree with them: an unproblematic state is a state without creative thought. Its other name is death.

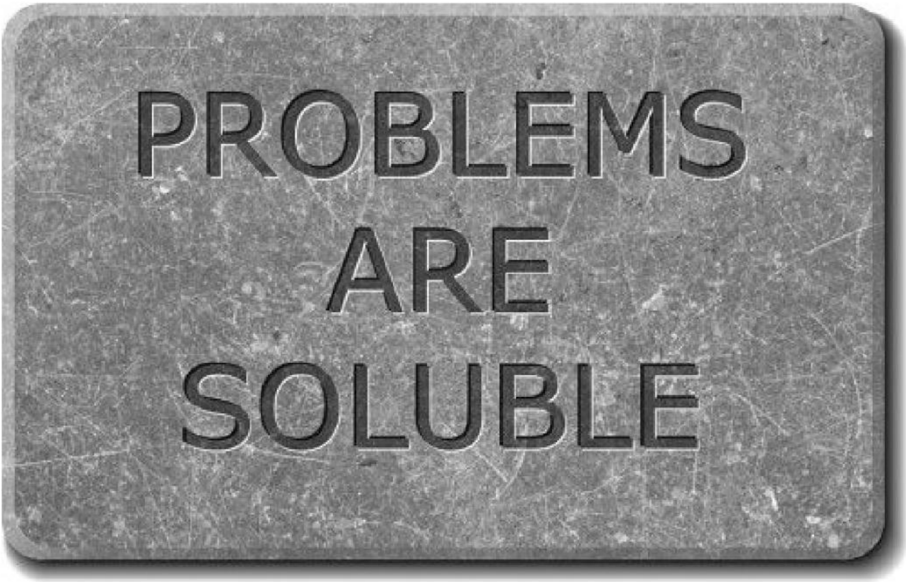
All those kinds of problem (survival-related, progress-related, moral, and sheer-curiosity-driven problems) are connected. We can, for instance, expect that our ability to cope with existential threats will continue to depend on knowledge that was originally created for its own sake. And we can expect disagreements about goals and values always to exist, because, among other reasons, moral explanations depend partly on facts about the physical world. For instance, the moral stances in the Principle of Mediocrity and the Spaceship Earth idea depend on the physical world not being explicable in the sense that I have argued it must be.

Nor will we ever run out of problems. The deeper an explanation is, the more new problems it creates. That must be so, if only because there can be no such thing as an ultimate explanation: just as 'the gods did it' is always a bad explanation, so any other purported foundation of all explanations must be bad too. It must be easily variable because it cannot answer the question: why that foundation and not another? Nothing can be explained only in terms of itself. That holds for philosophy just as it does for science, and in particular it holds for *moral* philosophy: no utopia is possible, but only because our values and our objectives can continue to improve indefinitely.

Thus fallibilism alone rather understates the error-prone nature of knowledge-creation. Knowledge-creation is not only *subject* to error: errors are common, and significant, and always will be, and correcting them will always reveal further and better problems. And so the maxim that I suggested should be carved in stone, namely 'The Earth's biosphere is *incapable* of supporting human life' is actually a special case of a much more general truth, namely that, for people, *problems are inevitable*. So let us carve *that* in stone:



It is inevitable that we face problems, but no particular problem is inevitable. We survive, and thrive, by solving each problem as it comes up. And, since the human ability to transform nature is limited only by the laws of physics, none of the endless stream of problems will ever constitute an impassable barrier. So a complementary and equally important truth about people and the physical world is that *problems are soluble.* By 'soluble' I mean that the right knowledge would solve them. It is not, of course, that we can possess knowledge just by wishing for it; but it is in principle accessible to us. So let us carve that in stone too:



That *progress* is both possible and desirable is perhaps the quintessential idea of the Enlightenment. It motivates all traditions of criticism, as well as the principle of seeking good explanations. But it can be interpreted in two almost opposite ways, both of which, confusingly, are known as 'perfectibility'. One is that humans, or human societies, are capable of attaining a state of supposed perfection – such as the Buddhist or Hindu 'nirvana', or various political utopias. The other is that every attainable state can be indefinitely improved. Fallibilism rules out that first position in favour of the second. Neither the human condition in particular nor our explanatory knowledge in general will ever be perfect, nor even approximately perfect. We shall always be at the *beginning* of infinity.

These two interpretations of human progress and perfectibility have historically inspired two broad branches of the Enlightenment which, though they share attributes such as their rejection of authority, are so different in important respects in that it is most unfortunate that they share the same name. The utopian 'Enlightenment' is sometimes called the Continental (European) Enlightenment to distinguish it from the more fallibilist British Enlightenment, which began a little earlier and took a very different course. (See, for instance, the historian Roy Porter's book *Enlightenment*.) In my terminology, the Continental Enlightenment understood that problems are soluble but not that they are inevitable, while the British Enlightenment understood both equally. Note that this is a classification of ideas, not of nations or even individual thinkers: not all Enlightenment thinkers belong wholly to one branch or the other; nor were all thinkers of the respective Enlightenments born in the eponymous part of the world. The mathematician and philosopher Nicholas de Condorcet, for instance, was French yet belonged more to what I am calling the 'British' Enlightenment, while Karl Popper, the twentieth century's foremost proponent of the British Enlightenment, was born in Austria.

The Continental Enlightenment was impatient for the perfected state – which led to intellectual dogmatism, political violence and new forms of tyranny. The French Revolution of 1789 and the Reign of Terror that followed it are the archetypal examples. The British Enlightenment, which was evolutionary and cognizant of human fallibility, was impatient for institutions that did not stifle gradual, continuing change. It was also enthusiastic for small improvements, unbounded in the future. (See, for instance, the historian Jenny Uglow's book *Lunar Men*.) This is, I believe, the