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 halfway across the known universe. It was probably a single star collapsing to form a black hole – an object whose gravity is so intense that not even light can escape from its interior. The explosion was intrinsically brighter than a million supernovae, and would have been visible with the naked eye from Earth – though only faintly and for only a few seconds, so it is unlikely that anyone here saw it. Supernovae last longer, typically fading on a timescale of months, which allowed astronomers to see a few in our galaxy even before the invention of telescopes.

Another class of cosmic monsters, the intensely luminous objects known as *quasars*, are in a different league. Too distant to be seen with the naked eye, they can outshine a supernova for millions of years at a time. They are powered by massive black holes at the centres of galaxies, into which entire

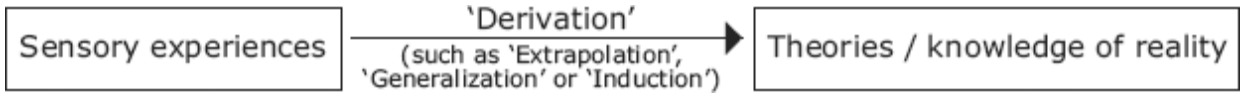
stars are falling – up to several per day for a large quasar – shredded by tidal effects as they spiral in. Intense magnetic fields channel some of the gravitational energy back out in the form of jets of high-energy particles, which illuminate the surrounding gas with the power of a trillion suns.

Conditions are still more extreme in the black hole's interior (within the surface of no return known as the 'event horizon'), where the very fabric of space and time may be being ripped apart. All this is happening in a relentlessly expanding universe that began about fourteen billion years ago with an all-encompassing explosion, the Big Bang, that makes all the other phenomena I have described seem mild and inconsequential by comparison. And that whole universe is just a sliver of an enormously larger entity, the multiverse, which includes vast numbers of such universes.

The physical world is not only much bigger and more violent than it once seemed, it is also immensely richer in detail, diversity and incident. Yet it all proceeds according to elegant laws of physics that we understand in some depth. I do not know which is more awesome: the phenomena themselves or the fact that we know so much about them.

How do we know? One of the most remarkable things about science is the contrast between the enormous reach and power of our best theories and the precarious, local means by which we create them. No human has ever been at the surface of a star, let alone visited the core where the transmutation happens and the energy is produced. Yet we see those cold dots in our sky and *know* that we are looking at the white-hot surfaces of distant nuclear furnaces. Physically, that experience consists of nothing other than our brains responding to electrical impulses from our eyes. And eyes can detect only light that is inside them at the time. The fact that the light was emitted very far away and long ago, and that much more was happening there than just the emission of light – those are not things that we see. We know them only from theory.

Scientific theories are *explanations*: assertions about what is out there and how it behaves. Where do these theories come from? For most of the history of science, it was mistakenly believed that we 'derive' them from the evidence of our senses – a philosophical doctrine known as *empiricism*:



Empiricism

For example, the philosopher John Locke wrote in 1689 that the mind is like 'white paper' on to which sensory experience writes, and that that is where all our knowledge of the physical world comes from. Another empiricist metaphor was that one could *read* knowledge from the 'Book of Nature' by making observations. 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 embedded in a hollow, rotating ‘celestial sphere’ centred on the Earth (or that they were holes in the sphere, through which the light of heaven shone). This *geocentric* – Earth-centred – theory of the universe seemed to have been directly derived from experience, and repeatedly confirmed: anyone who looked up could ‘directly observe’ the celestial sphere, and the stars maintaining their relative positions on it and being held up just as the theory predicts. Yet in reality, the solar system is *heliocentric* – centred on the sun, not the Earth – and the Earth is not at rest but in complex motion. Although we first noticed a daily rotation by observing stars, it is not a property of the stars at all, but of the Earth, and of the observers who rotate with it. It is a classic example of the deceptiveness of the senses: the Earth looks and feels as though it is at rest beneath our feet, even though it is really rotating. As for the celestial sphere, despite being visible in broad daylight (as the sky), it does not exist at all.

The deceptiveness of the senses was always a problem for empiricism – and thereby, it seemed, for science. The empiricists’ best defence was that the senses cannot be deceptive in themselves. What misleads us are only the false interpretations that we place on appearances. That is indeed true – but only because our senses themselves do not say anything. Only our interpretations of them do, and those are very fallible. But the real key to science is that our explanatory theories – which include those interpretations – can be *improved*, through conjecture, criticism and testing.

Empiricism never did achieve its aim of liberating science from authority. It denied the legitimacy of traditional authorities, and that was salutary. But unfortunately it did this by setting up two other false authorities: sensory experience and whatever fictitious process of ‘derivation’, such as induction, one imagines is used to extract theories from experience.

The misconception that knowledge needs authority to be genuine or reliable dates back to antiquity, and it still prevails. To this day, most courses in the philosophy of knowledge teach that knowledge is some form of *justified, true belief*, where ‘justified’ means designated as true (or at least ‘probable’) by reference to some authoritative source or touchstone of knowledge. Thus ‘how do we *know* . . . ?’ is transformed into ‘by what authority do we claim .

. . ?’ The latter question is a chimera that may well have wasted more philosophers’ time and effort than any other idea. It converts the quest for truth into a quest for certainty (a feeling) or for endorsement (a social status). This misconception is called *justificationism.*

The opposing position – namely the recognition that there are no authoritative sources of knowledge, nor any reliable means of justifying ideas as being true or probable – is called *fallibilism*. To believers in the justified-true-belief theory of knowledge, this recognition is the occasion for despair or cynicism, because to them it means that knowledge is unattainable. But to those of us for whom creating knowledge means understanding better what is really there, and how it really behaves and why, fallibilism is part of the very means by which this is achieved. Fallibilists expect even their best and most fundamental explanations to contain misconceptions in addition to truth, and so they are predisposed to try to change them for the better. In contrast, the logic of justificationism is to seek (and typically, to believe that one has found) ways of securing ideas *against* change. Moreover, the logic of fallibilism is that one not only seeks to correct the misconceptions of the past, but hopes in the future to find and change mistaken ideas that no one today questions or finds problematic. So it is fallibilism, not mere rejection of authority, 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 false, could be contradicted by the outcome of some possible observation. Thus, although scientific theories are not derived from experience, they can be tested by experience – by observation or experiment. For example, before the discovery of radioactivity, chemists had believed (and had verified in countless experiments) that transmutation is impossible. Rutherford and Soddy boldly conjectured that uranium spontaneously transmutes into other elements. Then, by demonstrating the creation of the element radium in a sealed container of uranium, they refuted the prevailing theory and science progressed. They were able to do that because that earlier theory was testable: it was possible to test for the presence of radium. In contrast, the ancient theory that all matter is composed of combinations of the elements earth, air, fire and water was untestable, because it did not include any way of testing for the presence of those components. So it could never be refuted by experiment. Hence it could never be – and never was – improved upon through experiment. The Enlightenment was at root a philosophical change.

The physicist Galileo Galilei was perhaps the first to understand the importance of experimental tests (which he called *cimenti*, meaning ‘trials by ordeal’) as distinct from other forms of experiment and observation, which can more easily be mistaken for ‘reading from the Book of Nature’.

Testability is now generally accepted as the defining characteristic of the scientific method. Popper called it the ‘criterion of demarcation’ between science and non-science.

Nevertheless, testability cannot have been the decisive factor in the scientific revolution either. Contrary to what is often said, testable predictions had always been quite common. Every traditional rule of thumb for making a flint blade or a camp fire is testable. Every would-be prophet who claims that the sun will go out next Tuesday has a testable theory. So does every gambler who has a hunch that ‘this is my lucky night – I can feel it’. So what is the vital, progress-enabling ingredient that is present in science, but absent from the testable theories of the prophet and the gambler?

The reason that testability is not enough is that prediction is not, and cannot be, the purpose of science. Consider an audience watching a conjuring trick. The problem facing them has much the same logic as a scientific problem. Although in nature there is no conjurer trying to deceive us intentionally, we can be mystified in both cases for essentially the same reason: appearances are not self-explanatory. If the explanation of a conjuring trick were evident in its appearance, there would be no trick. If the explanations of physical phenomena were evident in their appearance, empiricism would be true and there would be no need for science as we know it.

The problem is not to predict the trick’s appearance. I may, for instance, predict that if a conjurer seems to place various balls under various cups, those cups will later appear to be empty; and I may predict that if the conjurer appears to saw someone in half, that person will later appear on stage unharmed. Those are testable predictions. I may experience many conjuring shows and see my predictions vindicated every time. But that does not even address, let alone solve, the problem of how the trick works. Solving it requires an explanation: a statement of the reality that accounts for the appearance.

Some people may enjoy conjuring tricks without ever wanting to know how they work. Similarly, during the twentieth century, most philosophers, and many scientists, took the view that science is incapable of discovering anything about reality. Starting from empiricism, they drew the inevitable

conclusion (which would 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 common-sense, 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 cupsand-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 modifications of ideas that were not originally conjuring tricks – for instance, ideas for hiding objects in earnest. Similarly, where did the first scientific ideas come from? Before there was science there were rules of thumb, and explanatory assumptions, and myths. So there was plenty of raw material for criticism, conjecture and experiment to work with. Before that, there were our inborn assumptions and expectations: we are born with ideas,

and with the ability to make progress by changing them. And there were patterns of cultural behaviour – about which I shall say more in Chapter 15.

But even *testable, explanatory theories* cannot be the crucial ingredient that made the difference between no-progress and progress. For they, too, have always been common. Consider, for example, the ancient Greek myth for explaining the annual onset of winter. Long ago, Hades, god of the underworld, kidnapped and raped Persephone, goddess of spring. Then Persephone's mother, Demeter, goddess of the earth and agriculture, negotiated a contract for her daughter's release, which specified that Persephone would marry Hades and eat a magic seed that would compel her to visit him once a year thereafter. Whenever Persephone was away fulfilling this obligation, Demeter became sad and would command the world to become cold and bleak so that nothing could grow.

That myth, though comprehensively false, does constitute an explanation of seasons: it is a claim about the reality that brings about our experience of winter. It is also eminently testable: if the cause of winter is Demeter's periodic sadness, then winter must happen everywhere on Earth at the same time. Therefore, if the ancient Greeks had known that a warm growing season occurs in Australia at the very moment when, as they believed, Demeter is at her saddest, they could have inferred that there was something wrong with their explanation of seasons.

Yet, when myths were altered or superseded by other myths over the course of centuries, the new ones were almost never any closer to the truth. Why? Consider the role that the specific elements of the Persephone myth play in the explanation. For example, the gods provide the *power* to affect a large-scale phenomenon (Demeter to command the weather, and Hades and his magic seed to command Persephone and hence to affect Demeter). But why those gods and not others? In Nordic mythology, seasons are caused by the changing fortunes of Freyr, the god of spring, in his eternal war with the forces of cold and darkness. Whenever Freyr is winning, the Earth is warm; when he is losing, it is cold.

That myth accounts for the seasons about as well as the Persephone myth. It is slightly better at explaining the randomness of weather, but worse at explaining the regularity of seasons, because real wars do not ebb and flow

so regularly (except insofar as that is due to seasons themselves). In the Persephone myth, the role of the marriage contract and the magic seed is to explain that regularity. But why is it specifically a magic seed and not any other kind of magic? Why is it a conjugalvisits contract and not some other reason for someone to repeat an action annually? For instance, here is a variant explanation that fits the facts just as well: Persephone was not released – she escaped. Each year in spring, when her powers are at their height, she takes revenge on Hades by raiding the underworld and cooling all the caverns with spring air. The hot air thus displaced rises into the human world, causing summer. Demeter celebrates Persephone's revenge and the anniversary of her escape by commanding plants to grow and adorn the Earth. This myth accounts for the same observations as the original, and it is testable (and in fact refuted) 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 mythmaking 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 is it 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 different times of year. In the Persephone myth, in contrast, the coldness of the world is caused by Demeter's sadness – but people do not generally cool their surroundings when they are sad, and we have no way of knowing that Demeter *is* sad, or that she ever cools the world, other than the onset of winter itself. One could not substitute the moon for the sun in the axis-tilt story, because the position of the moon in the sky does not repeat itself once a year, and because the sun's rays heating the Earth are integral to the explanation. Nor could one easily incorporate any stories about how the sun god feels about all this, because if the true explanation of winter is in the geometry of the Earth–sun motion, then how anyone feels about it is irrelevant, and if there were some flaw in that explanation, then no story about how anyone felt would put it right.

The axis-tilt theory also predicts that the seasons will be out of phase in the two hemispheres. So if they had been found to be in phase, the theory would have been refuted, just as, in the event, the Persephone and Freyr myths were refuted by the opposite observation. But the difference is, if the axis-tilt theory had been refuted, its defenders would have had nowhere to go. No easily implemented change could make tilted axes cause the same seasons all over the planet. Fundamentally new ideas would have been needed. That is what makes good explanations essential to science: it is only when a theory is a good explanation – hard to vary – that it even matters whether it is testable. Bad explanations are equally useless whether they are testable or not.

Most accounts of the differences between myth and science make too much of the issue of testability – as if the ancient Greeks' great mistake was that they did not send expeditions to the southern hemisphere to observe the seasons. But in fact they could never have guessed that such an expedition

might provide evidence about seasons unless they had already guessed that seasons would be out of phase in the two hemispheres – and if that guess was hard to vary, which it could have been only if it had been part of a good explanation. If their guess was *easy* to vary, they might just as well have saved themselves the boat fare, stayed at home, and tested the easily testable theory that winter can be staved off by yodelling.

So long as they had no better explanation than the Persephone myth, there should have been no need for testing. Had they been seeking good explanations, they would immediately have tried to improve upon the myth, without testing it. That is what we do today. We do not test every testable theory, but only the few that we find are good explanations. Science would be impossible if it were not for the fact that the overwhelming majority of false theories can be rejected out of hand without any experiment, simply for being bad explanations.

Good explanations are often strikingly simple or elegant – as I shall discuss in Chapter 14. Also, a common way in which an explanation can be bad is by containing superfluous features or arbitrariness, and sometimes removing those yields a good explanation. This has given rise to a misconception known as 'Occam's razor' (named after the fourteenth-century philosopher William of Occam, but dating back to antiquity), namely that one should always seek the 'simplest explanation'. One statement of it is 'Do not multiply assumptions beyond necessity.' However, there are plenty of very simple explanations that are nevertheless easily variable (such as 'Demeter did it'). And, while assumptions 'beyond necessity' make a theory bad by definition, there have been many mistaken ideas of what is 'necessary' in a theory. Instrumentalism, for instance, considers explanation itself unnecessary, and so do many other bad philosophies of science, as I shall discuss in 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 bounded? Or is this the beginning of infinity – that is to say, do these methods have unlimited potential to create further knowledge? It may seem paradoxical to claim anything so grand (even if only potentially) on behalf of a project that has swept away all the ancient myths that used to assign human beings a special significance in the scheme of things. For if the power of the human faculties of reason and creativity, which have driven the Enlightenment, were indeed unlimited, would humans not have just such a significance?

And yet, as I mentioned at the beginning of this chapter, gold can be created only by stars and by intelligent beings. If you find a nugget of gold anywhere in the universe, you can be sure that in its history there was either a supernova or an intelligent being with an explanation. And if you find an explanation anywhere in the universe, you know that there must have been an intelligent being. A supernova alone would not suffice.

But – so what? Gold is important *to us*, but in the cosmic scheme of things it has little significance. Explanations are important to us: we need them to survive. But is there anything significant, in the cosmic scheme of things, about explanation, that apparently puny physical process that happens inside brains? I shall address that question in Chapter 3, after some reflections about appearance and reality.

TERMINOLOGY

Explanation Statement about what is there, what it does, and how and why.

Reach The ability of some explanations to solve problems beyond those that they were created to solve.

Creativity The capacity to create new explanations.

Empiricism The misconception that we 'derive' all our knowledge from sensory experience.

Theory-laden There is no such thing as 'raw' experience. All our experience of the world comes through layers of conscious and unconscious interpretation.

Inductivism The misconception that scientific theories are obtained by generalizing or extrapolating repeated experiences, and that the more often a theory is confirmed by observation the more likely it becomes.

Induction The non-existent process of 'obtaining' referred to above.

Principle of induction The idea that 'the future will resemble the past', combined with the misconception that this asserts anything about the future.

Realism The idea that the physical world exists in reality, and that knowledge of it can exist too.

Relativism The misconception that statements cannot be objectively true or false, but can be judged only relative to some cultural or other arbitrary standard.

Instrumentalism The misconception that science cannot describe reality, only predict outcomes of observations.

Justificationism The misconception that knowledge can be genuine or reliable only if it is justified by some source or criterion.

Fallibilism The recognition that there are no authoritative sources of knowledge, nor any reliable means of justifying knowledge as true or probable.

Background knowledge Familiar and currently uncontroversial knowledge.

Rule of thumb 'Purely predictive theory' (theory whose explanatory content is all background knowledge).

Problem A problem exists when a conflict between ideas is 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.