VINTAGE

CONTENTS

ABOUT THE AUTHOR

Julian Barbour is a former Visiting Professor in Physics at the University of Oxford and author of the highly regarded The Discovery of Dynamics and the bestseller The End of Time. His papers have been published in the world's most prestigious scientific journals, including Nature, Proceedings of the Royal Society and Physical Review Letters and he has made numerous appearances on national radio, television and in various documentaries. The Janus Point is his first book in twenty years and the culmination of five decades' work.

In memory of my wife, Verena, and our daughter Jessica

LIST OF ILLUSTRATIONS

PRAISE FOR *THE JANUS POINT*

"Julian Barbour's *The Janus Point* is simply the most important book I have read on cosmology in several years. He presents a novel approach to the central question of why time has a direction, providing a serious alternative to contemporary thinking. With a rare humanity and a perspective based on a lifetime of study of the history and philosophy of cosmology, Barbour writes a book that is both a work of literature and a masterpiece of scientific thought."

—LEE SMOLIN, author of *The Trouble with Physics*

"Julian Barbour is well known for his brilliant study of physics history, *The Discovery of Dynamics. The Janus Point* includes a similar history of thermodynamics, statistical mechanics, and the arrow of time. But for me the main point of the book was to show history-in-the-making. His 'shape dynamics' is a project to recast the foundations of all of cosmology, gravity, thermodynamics, and the arrow of time. The book has given me a lot to ponder. As Gauss said of Riemann's habilitation lecture, '[it] exceeded my expectations.'"

—BILL UNRUH, professor of physics at University of British Columbia

"Julian Barbour has no peer when it comes to explaining scientific ideas in a way that is accessible without being simplistic. For good measure he has a talent for using quotes from Shakespeare and other literary sources in a manner that actually helps to elucidate key points. In *The Janus Point* he tackles subject matter that is notoriously challenging even to scientists, and explains it in a way that gave me new insights and understanding even though I studied these topics in a classroom a long time ago. This is a fitting sequel to his earlier work and helps to pull together several big ideas that some of us have been watching with fascination for decades."

—NEAL STEPHENSON, author of *Snow Crash*

"By abandoning the prejudice that particles (atoms, stars ...) are confined in a box, Julian Barbour has discovered an unexpected and remarkably simple feature of Newtonian dynamics. It is the basis of his seductive and eloquently presented explanation of the history of the universe, even time itself. Is his cosmology correct? 'Time' will tell."

—MICHAEL VICTOR BERRY, Melville Wills Professor of Physics (Emeritus) at Bristol University

"Julian Barbour's infectious enthusiasm for the big ideas in physics is addictive. He has a complete mastery of the history of ideas yet a remarkable lightness and clarity in explaining what are profound concepts. *The Janus Point* is controversial and gripping, an extraordinary introduction to his view of the universe."

—PEDRO G. FERREIRA, author of *The Perfect Theory*

“Julian Barbour is a profound and original thinker, with the boldness to tackle some of nature’s deepest problems. He is also a fine writer, and this renders his book—despite its conceptual depth—accessible to anyone who has pondered the mysteries of space and time. It’s a distillation of the author’s prolonged investigations, and the insights that he offers deserve wide readership.”

—MARTIN REES, author of *On the Future*

PREFACE AND ACKNOWLEDGEMENTS

THIS BOOK'S ORIGINAL SUBTITLE was *A New Theory of Time's Arrows and the Big Bang*. That is what its substance remains, but I was happy to accept the suggestion of TJ Kelleher, my editor of the US edition, and adopt its present shorter form. I think there is warrant for it. The big bang not only gave birth to time but also stamped on it the eternal aspect of an arrow's flight. Thus the two together do amount to a new theory of time. Please note the indefinite article in the subtitle. Nothing in science is definitive. Hypotheses are proposed and tested. Science progresses when, through precise experiment and good observation, predictions are either confirmed or refuted. I believe the proposed explanation of time's arrows is as secure as is the long-established expansion of the universe, but some radical ideas about the big bang, which matured late in the writing of the book, are definitely speculative. I have nevertheless decided to include them because they do represent what seems to me, now that they have been recognised, to be an almost inevitable bringing together of everything else.

This is a very personal book in which I have tried to combine established science of the cosmos with new ideas, but I also include here and there my own reactions to existence in the universe and wonder at its nature. How is it that time has not only created the physical world of atoms and galaxies but also poets, painters, and composers? The works of Shakespeare have been a great joy in my life. You will find him quoted explicitly in a dozen or so places, but for buffs of the Bard I have also, now and then and unattributed, smuggled in from his plays and sonnets a half line or even a single word. I hope you will get a little pleasure if you spot these purloined feathers. The final chapter is not quite an epilogue because it brings in discussion of the arrow we experience most directly, that of the passage of time. I think this is intimately related to the greatest mystery of all—the gift of consciousness. Don't expect any answer to the mystery, but perhaps I can offer illumination of one of its aspects. Otherwise the chapter is an attempt to identify what it is in the mathematics of the universe that is manifested in art. It must be there since all great art has a unique structure, and structure is the very essence of mathematics.

As regards mathematics itself, inclusion of some in the book is inevitable, if only to avoid endless circumlocution. It is the concepts that count. Formulas are given for three concepts; that's all there are. Their names then make many appearances. The mere repetition may help. The renowned mathematician John von Neumann is reputed to have said, "Young man, you do not understand mathematics, you get used to it". There is a lot of truth in that. I think all my readers know perfectly well what the circumference of a circle is. With luck you will come to a similar intuition for the single most important concept in the book. I call it *complexity*. As I do here, I have used italics in the book almost exclusively to flag the first appearance of an important concept.

At the end of the book there is a list of the figures—there are twenty-six—and the page on which you will find them. Also at the end of the book, together with some additional material, I have compiled some technical notes for readers with at least some background in physics and mathematics, roughly from first-year university students to advanced researchers who want to see the evidence for statements in the body of the book and perhaps follow them up. There is also a bibliography restricted to books and technical papers that have a more or less direct relation to this one.

It includes my own *The End of Time*, published in 1999. I mention it here because some of the readers of this book will have read it and may wonder how my second venture into the genre of books for general readers differs from it. All I will say here is that the two books cover the same theme—time's arrows and their origin—but from a point of view that is somewhat different. When appropriate, I draw attention to specific differences in footnotes. There is, of course, an index.

This book could never have appeared or taken the form it does without critical input during the last eight years from my current collaborators Tim Koslowski, Flavio Mercati, and David Sloan. Some of the most important ideas come from them. It has been a great pleasure to work with them—Tim since we first met at the Perimeter Institute in Canada in 2008, Flavio since 2011, and Dave since 2015. Flavio also generated all but one of the book's figures and found online the one he did not create—the fine double-headed Janus on a Roman coin in Fig. 3. The figures are a very important part of the book and I am most grateful to Flavio for them.

To continue with acknowledgements: What we call *shape dynamics*—it's a representation of the dynamical essence of the theories of both Newton and Einstein—forms the mathematical core of this book; recent developments of it owe much to Tim, Flavio, and Dave. The notes include references to people who either directly or indirectly contributed to its earlier development, but I should mention here Bruno Bertotti (with whom I collaborated closely from the mid-1970s until 1982 and whom I was able to see in September 2018, just a month before he died), Niall Ó Murchadha, Bryan Kelleher, Brendan Foster, Edward Anderson, Sean Gryb (whom I also met at the Perimeter Institute through my friend Lee Smolin, who was his PhD supervisor there), and Henrique Gomes. Numerous visits to the Perimeter Institute, which hosted three workshops on shape dynamics, were much appreciated, as were, at the invitation of Viqar Husain, two visits (one for a workshop) at the University of New Brunswick in Fredericton. Sean Gryb also organised a valuable shape dynamics workshop at Nijmegen in the Netherlands and, with Karim Thébault, two at the University of Bristol. Two other friends of long standing with whom I have had many valuable and informative discussions are Karel Kuchař, of the University of Utah in Salt Lake City, and Christopher Isham, of Imperial College in London.

Although they are not involved in shape dynamics, I must here also thank very warmly Alain Chenciner and Alain Albouy, of the Institut de Mécanique Céleste et de Calcul des Éphémérides at the Observatory of Paris. Through the kind introduction of Richard Montgomery, of the Department of Mathematics at the University of California, Santa Cruz, I have been interacting with them on and off for twenty years. With their help I have learned much about the oldest and still in many ways the most important problem in mathematical physics; it goes right back to Isaac Newton and concerns the behaviour of point particles that interact in accordance with Newton's laws of motion and universal gravitation. It turned out that some of the most beautiful results gleaned in its study, literally over centuries, are almost tailor-made to be the foundation of much of this book. It would not be what it is without their help. To combine visits to Paris with discussion at the venerable Observatory has been a rare pleasure.

Closer to my home, Harvey Brown, working in the philosophy of science at the University of Oxford, has been a good friend and sounding board for ideas in discussions over many years; I have also had useful discussions with his colleagues Simon Saunders and Oliver Pooley. Pedro Ferreira, another professor at Oxford, in the Department of Astrophysics, has, almost inadvertently but through welcome interest in shape dynamics, played a critical role in determining the present content of my book. About five years ago he set Tim and me a challenge to see if, in the framework of Newtonian gravity represented in shape-dynamic form, we could find an alternative to what is called inflation in cosmology and is the basis of the current theory of large-scale structure formation in the universe. Pedro was also a great help in introducing us to the cosmologist Michael Joyce, who is also in Paris, though not at the Observatory. I am not going to claim we have met Pedro's challenge, but through it some entirely unexpected possibilities have come to light and figure prominently in the book. They relate to the nature of the big bang and, presented in Chapters 16 through 18, are the radical ideas mentioned above. Because they may be controversial, this is where, as the author, I must stick my colours to the mast. Hypotheses, especially if they have some plausibility, are, I think, acceptable; outright errors are quite another thing. In this book I am, especially in the ideas about the big bang and the final state of the universe, going out on something of a limb. If it breaks or I fall, any and all mistakes in the book are entirely my responsibility.

I must also warmly thank my agent, Max Brockman, of Brockman Inc in New York, and his parents, John Brockman and Katinka Matson, who handled my book *The End of Time*. I also want to very especially thank my two editors: TJ Kelleher for Basic Books in the United States (he tells me TJ is the invariable moniker by which he is known, both to family and to colleagues) and Will Hammond of Bodley Head in the UK. Having been involved with few books dealing with science for general readers, Will was glad to leave editing to TJ. Work with him has been most stimulating. He has been just what an editor should be: supportive throughout but clear about what must go or be changed.[fn1]

One major excision I will mention. I began work on this book three and a half years ago thinking I needed not only to study the history of thermodynamics, the discovery and development of which first brought to the fore the mystery of time's arrows, but also to include in the book a lot of the history, which has certainly given me a strong sense of its importance. As TJ pointed out, the first draft contained far too much; all that now remains explicitly is the minimum needed to set the scene and introduce some of the concepts and issues that feature in the heart of the book. However, there remains an arc of history that still informs the structure of the book. Moreover, several friends and colleagues were gratifyingly positive about the history and said it should see the light of day. I have therefore taken the chapters that were written but have been removed and put them as a PDF on my website, platonia.com; I hope to be able in the not too distant future to tidy them up as a book on the history of thermodynamics. You get that only in a very condensed form in this book; I hope anyone interested in the history will visit my website and perhaps download the PDF.

That brings me to one more thing: the great value of the internet and especially Wikipedia. When I next see an appeal to support Wikipedia, I do intend to make a contribution. I also recommend you check out Google Images for images of the various scientists I mention and look at Wikipedia for their biographies. It is all so easy nowadays; I think you will find it well worth the minor effort of a click.

Mention of modern digital technology brings me to thank Melissa Veronesi of Basic Books for her help with final checking of the text as it came to me from the copy editor, Sue Warga, who has done a great job. It has been an eye-opener to see in how many ways text can be made to flow better. I am also very grateful to my son, Boris, for help with digital issues and preparation of the text.

Boris is not the only member of my family I want to thank. My wife, Verena, succumbed to Alzheimer's just after I submitted the proposal for this book to my agent in June 2016; she had remained wonderfully cheerful to the end. But in the midst of the writing of this book, eighteen months after her mother's death, I lost my daughter Jessica. Her two sisters, Naomi and Dorcas, together with Boris, have been a great comfort, as have eight grandchildren, including Jessica's two daughters. You will see that the book is dedicated to the memory of Verena and Jessica. Both very greatly enriched my life.

CHAPTER 1

TIME AND ITS ARROWS

The Universe is made of stories, not of atoms.

—MURIEL RUKEYSER

TIME FLOWS FORWARD. EVERYONE HAS THE FEELING. IT IS MORE THAN MERE feeling; it has a real counterpart in observable phenomena. Processes near and far in space and time all unfold in the same direction. All animals, us humans included, get older in the same direction. We never meet anyone getting younger. Astronomers have observed millions of stars and understand very well how they age—all in the same direction as us. On seashores around the world, waves build and break—they never 'unbreak'.

We remember the past, not the future. There are arrows of time. A film run backwards confounds cause and effect: instead of divers making a splash on entering the water, they emerge from it while the splash disappears. Myriad arrows permeate our existence. They are the stuff of birth, life, and death. In their totality, they define for us the direction of time.

Three arrows are particularly important because they can be treated with a good degree of mathematical precision.

The first, with which we will be much concerned, is the common process of equilibration. To see an example and its end result, put a tumbler of water on a table and disturb its surface by stirring the water vigorously with your finger. Remove your finger. Very soon the disturbance subsides and the surface becomes flat. You have witnessed an irreversible process. You can watch the surface for hours on end and it will never become disturbed spontaneously. The observationally unchanging state reached through equilibration corresponds to equilibrium. This is an example you can see. More important for the subject of this book is one you can feel— the equalisation of temperature. Go from a warm room to a cold room and you immediately feel the difference because your body is losing heat to the air around it. Many similar examples could be given: hot coffee in a mug, if not drunk, cools to the ambient temperature of the room.

The next arrow relates to what are called retarded waves. Don't worry about the name. You see a beautiful example whenever you throw a stone into a still pond and circular waves spread out from the point of entry. The waves are said to be retarded because they are observed after the impact of the stone—the effect follows the cause. You never see waves that mysteriously start in unison at the bank of the pond, converge on its centre, and eject a stone that then lands in your hand—although the laws of hydrodynamics and mechanics are perfectly compatible with the possibility. It is not only water waves that invariably exhibit retardation; so do the radio and TV signals that reach your home from transmitters.

The third example is in quantum mechanics and is the notorious problem of Schrödinger's cat. In accordance with the quantum formalism, a certain wave function can describe a cat that is simultaneously both dead and alive. Only when an observation is made to establish the cat's state is there collapse of the wave function and just one of the two possibilities is realised.

Arthur Eddington, the British astrophysicist who in 1919 made Einstein into a world celebrity overnight with a famous telegramme to the London *Times* that confirmed a prediction of the general theory of relativity, coined the expression 'arrow of time' in the 1920s. Eddington had in mind mainly the arrow associated with equilibration, but people often use it as a portmanteau expression to cover all the arrows, as I will.

BECAUSE THE ARROWS of time are so ubiquitous it's easy to take them for granted, but since the early 1850s theoreticians have seen in them a major problem. It is easily stated: apart from a tiny temporal asymmetry in one single law that, just after the big bang, was one of the factors that prevented complete mutual annihilation of matter and antimatter but cannot have played any significant role in creating the currently observed huge asymmetry between past and future, all the remaining laws of nature make no distinction between past and future. They work equally well in both time directions. Take a very simple law, the one that governs billiard balls. Unlike divers plunging into water, a film that spans their impact does look the same forwards and backwards. Physicists say laws like that are time-reversal symmetric. By this, they do not mean that time is reversed but that if at some stage all relevant velocities are precisely reversed, the balls will retrace, at the same speeds, the paths previously taken. More complicated is the case of charged particles in a magnetic field; the direction of the field must also be reversed along with the particle velocities if the retracing is to occur. There is an even more subtle case related to electric charges, mirror reflection, and the single fortuitous rule-breaking exception which allowed a minuscule amount of matter to survive after the big bang and we who are made of it to come into existence billions of years later. In all cases, the problem is the same: if the individual laws of nature that count do not distinguish a direction of time, how does it come about that innumerable phenomena do?

This question throws up such basic issues that we cannot hope to answer it unless we identify secure foundations for our theorising. One such foundation is that all meaningful statements in science are about relationships. This applies to the very notion of the direction of time. We recognise one because all around us we have those multitudinous unidirectional arrows. Without their constant presence, a single diver emerging backwards out of turbulent water, leaving it smooth and flat, would not appear to violate the normal course of nature. Also important are the nature and precision of observation. Wearing night-vision spectacles sensitive to infrared light, we would not see the collision of billiard balls as the same backwards and forwards—we would see hot spots appear on both balls after the collision, while the film run backwards would show those spots disappearing. In cricket, infrared imaging cameras are used when umpires are in doubt whether the ball has nicked the bat or pad and, through friction, raised the temperature at a localised spot.

This sensitivity to the means of observation raises a critical question: are the laws of nature truly time-reversal symmetric? The answer almost universally given is that yes, they are at the fundamental level of elementary particles—there is no microscopic arrow of time, only one at the macroscopic level. This was the conclusion that scientists reached in the second half of the nineteenth century. It came about in the first place through one truly remarkable study. Towards the end of the eighteenth century, people started to investigate seriously the properties of heat. A large part of the stimulus for this was to understand the workings of steam engines and how to make them as efficient as possible. In 1824, the young French engineer Sadi Carnot published a book, remarkable for its profundity and brevity, on this problem. Initially it passed unnoticed, but in 1849 a paper by William Thomson (later Lord Kelvin) brought Carnot's work to the notice of Rudolf Clausius, with dramatic effect: together with Thomson, Clausius played a key role in the creation of an entirely new science, for which Thomson coined the name thermodynamics.

Its first law states that energy can be neither created nor destroyed. Within physics, this formalises the long-held belief reflected in Lear's warning to Cordelia: "Nothing will come of nothing". The second law of thermodynamics introduces the fundamental concept of entropy, a great discovery by Clausius, who introduced the name. Entropy is, to say the least, a subtle notion that, since Clausius's discovery, has been formulated in many different ways. Indeed, a search through the technical literature throws up about twenty different definitions of the second law. Some of them include entropy; others do not. Despite that, entropy is one of those scientific concepts that have entered common parlance along with energy, the big bang, the expansion of the universe, black holes, evolution, DNA, and a few more. In a non-technical simplification that can mislead, entropy is widely described as a measure of disorder. The most common formulation of the second law states that, under controlled conditions that allow proper definition, entropy cannot decrease and generally will increase. In particular, entropy always increases in any process of equilibration in a confined space. Among the various arrows of time, entropy's growth is the one that the majority of scientists take to be the most fundamental. The second law, often referred to without the addition 'of thermodynamics', has acquired an almost inviolable aura. In quite large part this is due to Clausius, who knew how to make sure that, deservedly, his results lodged securely in the mind. The paper of 1865 in which he coined 'entropy' ends with the words "Die Entropie der Welt strebt einem Maximum zu" (the entropy of the universe tends to a maximum). This statement had the impact that Clausius intended and was taken to mean that the universe will eventually expire in the heat death of thermal equilibrium. In human terms, there's a near anticipation of this in the final quatrain of Shakespeare's Sonnet 73:

> In me thou see'st the glowing of such fire,
> That on the ashes of his youth doth lie,
> As the death-bed whereon it must expire,
> Consum'd with that which it was nourish'd by.

Not surprisingly, the second law is grist for the mill of all pessimists. For many scientists, the growth of entropy, and with it disorder, is the ineluctable arrow that puts the direction into time. The mystery then is how the arrow gets into things so profoundly if the laws give no indication that it should.

THE DIFFICULTY RESIDES in the structure of the laws. By themselves they do not tell you what will happen in any given situation. You have to specify an initial condition. In the billiard example, you need to say where the two balls are at an initial time and the velocities they have at that moment. The law will then tell you what subsequently happens. It will give you a solution. The situation is this: *Law + Initial Condition → Solution*. In the simplest billiard example, the solution is time-reversal symmetric, just like the law that creates it. But the examples of processes with which I began this book most definitely are not. As I indicated then, scientists have not yet been able to resolve the mismatch between the symmetric laws and the asymmetric solutions.

What they have been able to do is at best give a partial solution. To illustrate this, one needs a game with more balls than billiards: snooker. The game begins with the white cue ball striking the triangle of fifteen reds. Successive impacts send the balls flying in all directions. No individual two-ball impact distinguishes a direction of time. But what then happens does. That's where the law which at the elementary two-ball (microscopic) level describes a phenomenon without an arrow is able to create a many-ball effect with a macroscopic arrow. Theoreticians are universally in agreement that many dynamical 'agents' (the balls in snooker, elementary particles in fundamental physics) must be present if a macroscopic arrow is to emerge from time-symmetric microscopic laws.

However, this is at best a necessary condition; by itself it is not sufficient. To understand this, make a film of the initial impact in a game of snooker, its explosive effect, and a few bounces off the table walls. Run it backwards. Each individual impact again satisfies the reversible billiard law. But, in a seeming miracle, reds with apparently random speeds and directions of motion conspire in the midst of their chaos to come to rest in a perfect triangle and eject the white.

Of course, there is no miracle. The game begins with a special initial condition. That singles out a special solution. Only the tiniest fraction of all possible solutions is like that. But the example does at least show that reversible laws are compatible with solutions like 'unbreaking' waves, provided that sufficiently many objects are involved. There is no absolutely irreconcilable conflict with the laws of nature even if one has to invoke something close to divine intervention. De facto, this is what physicists are forced to do. They have to assume that at some time in the distant past, most probably at the big bang, the universe was in a special state of extraordinarily high order, the ultimate origin of all time's arrows.

The philosopher of science David Albert has dubbed this assumption the 'past hypothesis'. It was first proposed by the great Austrian physicist Ludwig Boltzmann in the 1890s as one possible way to explain what is now called the arrow of time. But it's a stopgap he did not favour because it does not respect temporal symmetry. Science aims to explain phenomena by laws, not by inexplicable initial conditions arbitrarily imposed. By that criterion, the hypothesis fails, as I think Albert himself would not deny. But this leaves us in a very uncomfortable place: the most profound aspects of existence are attributed not to law but to a special condition 'put in by hand'. It's not a resolution; it's an admission of defeat. However, I would not be writing this book if that were all I had to say. Plenty of other writers have already done a good job of describing the problem; their books are included in the bibliography and there are brief comments about them in the notes to this chapter.

INSTEAD, I'M GOING to suggest that the problem could have a genuine—and surprisingly simple—resolution. My collaborators and I stumbled on it a few years ago while working on a different problem. In this chapter I will give you an outline of our proposal and then fill out the necessary details in the chapters that follow. If the proposals that I present are on the right lines, I think that, taken together, they do amount to a new theory of time itself and not just its arrows. While the arrows by themselves represent a major aspect of time and a huge part of our deepest and most intimate experience, a proper understanding of their nature is impossible without a radical transformation of our notion of time. Accordingly, this book has two parts. The first sets the scene with a brief history of thermodynamics, formulates its principles, and explains why they fail to solve the problem of time's arrows. The second presents the proposed solution; it takes up the bulk of the book.

We need the history because thermodynamics grew out of a specific problem which arose at a particular point in time: the industrial revolution and Sadi Carnot's search for the most efficient way to operate steam engines. In that slim booklet published in 1824 he laid down principles of remarkable robustness—they have stood now for almost two hundred years—but they all apply in a 'box scenario', that is, to steam, gas, or any fluid in an impermeable and insulating container. Despite this critical condition, it is widely assumed that the laws of thermodynamics and the notion of entropy can be carried over more or less unchanged to cosmology. But the universe is not in a box; it is expanding, seemingly without impediment.

The steam engine cylinder was the fruitful womb of thermodynamics, not of time's arrows. The mother and birth pangs of the latter must be sought in the universe, in which principles of far greater power than those of steam engines hold sway. Carnot sought the most efficient use of coal, not its ultimate origin. Of course, he lived too early for that; we are better placed to take the principles of thermodynamics out of their box. For we know something which neither Carnot, Boltzmann, nor any other scientist in the nineteenth century had any inkling of: that the universe is expanding. That changes everything. The box-scenario principles must, without loss of power, be turned almost upside down to describe a universe that can expand freely. Only then can the mystery of time's arrows be solved.

In the next two chapters I will describe the history of thermodynamics and explain, as simply as I can, the original form of its principles so that in the new context their essential core can, despite the radical inversion, be seen to be intact and to suggest possibilities of which Boltzmann never could have dreamed.

The Copernican revolution is the prime example of the knock-on consequences a change of perspective can unleash. The great comet of 1577 led to one of the first, and it shows what can happen if a conceptual box is swept away. Around 1600 Johannes Kepler challenged the notion that crystal spheres carried the planets, arguing that Tycho Brahe's observations proved the comet to be so far away that if the crystal spheres existed, it would have had to crash into them. Hence the spheres could not exist. That prompted Kepler to comment with far-reaching effect: "From now on", he said, "the planets must find their way through the void like the birds through the air. We must philosophise about these things differently".

He did, and it led him to his laws of planetary motion, without which Isaac Newton never could have found his law of universal gravitation. That in turn set the scientific revolution on its as yet unstoppable course. The analogy between a physical (or conceptual) box and crystal spheres is not perfect, but it will do. Both impose constraints. Freed from the spheres, Kepler's planets could go their own way; freed from a box, atoms can fly off in all directions.

Copernicus's and Kepler's examples also show how even though a conceptual framework may be overthrown, its technical achievements may live on in a new guise. Copernicus did not introduce any new mathematical techniques when he set the earth in motion through the heavens. He took all the techniques he used from Ptolemy's *Almagest*, written about fourteen hundred years earlier. Kepler too used those very same methods time and again. His discoveries are unthinkable without them.

Thus, we too can 'think outside the box' while still using what was learned within it, suitably modified. Luckily, this can be done with a model barely more complicated than the simple ideal-gas model that was used, with great success, to give an atomistic explanation of gas thermodynamics and led to the discovery of entropy. Little more needs to be done than adapt it to a system like the universe, which current observations suggest will expand forever. This leads to the two major changes to current thinking that I propose.

THE FIRST, AN idea that came almost by happenstance, is that the big bang ceases to be an explosive birth of the universe and with it time. Instead it is a special point on the timeline of the universe, on either side of which the universe's size increases. I call this special location the Janus point—hence the title of the book. Everything related to this idea is based on long-known physics but, critically, interpreted in terms of ratios. The reason this helps is that general relativity, as usually interpreted, is said to break down precisely at the instant at which the universe comes into existence at the big bang. As we look back in time to that moment, Einstein's theory predicts that the size of the universe becomes zero and that matter densities and other quantities become infinite. Physicists take such infinities as a sign that something must be wrong with the theory. It is said to predict its own demise.

However, at any instant, the universe has both a size and a shape, and the latter is defined in terms of ratios. The difference is important. A triangle gives a simple illustration of the distinction between size and shape. From the lengths of the three sides, mathematicians form ratios that determine the internal angles of the triangle, which fix its shape. If you hold up a cardboard triangle in front of you with its surface at right angles to your line of sight and move the triangle toward and away from your eyes, I think you will agree that the shape is what truly characterises the triangle's intrinsic nature.[fn1] The size is relative, and it appears to change only because your external viewpoint does. Ants on the triangle could only observe how its size changes relative to them and determine how the 'universe' they form with it changes. Similarly in cosmology, only the variables which describe shape are observable and hence physical. Astronomers do not see the universe expanding; they see it changing its shape and from that deduce its expansion.

I need to say here that Janus points of two different kinds can exist: either the size of the universe becomes exactly zero or it merely passes through a minimal value. The latter situation, in which the shape variables can be continued smoothly and uniquely through the Janus point to 'another universe' on the other side, is already fully adequate to explain the arrow of time, and it gives rise to many striking and interesting effects. Most of the chapters devoted to the Janus-point idea describe those effects. The second possibility, that the size goes to zero exactly, has consequences that potentially are considerably more far-reaching and suggest a different kind of Janus point, a veritable big bang with multiple universes on either side of it. Moreover, they lead to a state at the Janus point that looks suspiciously like a past hypothesis of the kind I just said was only a stopgap. There is, however, a difference. First, it is, so far as I know, the first explicit example given of such a state; second, its very form and the manner in which it arises suggest that it is not 'put in by hand' but arises as a necessary consequence of a *law of the universe* that is more fundamental than the individual laws of nature and indeed from which they all emerge as approximate descriptions of locally observed phenomena. Unfortunately, the study of the zero-size case involves mathematical issues that are far from trivial, and I devote to it less space, in part at least because I cannot go beyond conjectures. However, those conjectures are of such interest that I do believe they deserve inclusion.

The second change to current thinking, prompted by the first, is the introduction of an entropy-like quantity—called entaxy, to distinguish it from conventional entropy defined for boxed systems— suitable for an 'unboxed universe'. In contrast to ordinary entropy, it does not increase but decreases. This does not conflict with the second law of thermodynamics, which—as proved by Clausius— applies only to confined systems within the universe. The decrease of this new quantity as the universe evolves reflects above all the growth not of disorder but of *complexity*, manifested as the birth and growth of structures and with them the formation of previously nonexistent subsystems that become effectively self-confined. It is within such subsystems that conventional entropy increases. In fact, complexity, defined by a formal mathematical expression, is a more important notion than entaxy; it reflects what we see in the universe and makes it possible to define entaxy.

Because of the way the structure in the universe, represented as complexity, evolves on the two sides of the Janus point, the notion of time is transformed. It no longer has one direction, from past to future, but instead has two: from a common past at the Janus point to two futures in the two directions away from it. Two *back-to-back* half-solutions, each a 'half-Janus', describe the universe. A divinity able to survey the whole of the universe's timeline would see the growth of its complexity, manifested in the formation of localised structures—the subsystems just mentioned—in the same two directions.

If growth of structure defines the direction of time, the divinity cannot fail to see it flip at the Janus point. From that special location, she could, like Janus, look in the two opposite directions of time at once. She would see the same kinds of thing happening both ways, above all structure being created out of near uniformity. The various arrows of time that I described at the start of this chapter would all be there, in both cases as if in flight from Janus. She would see that, in its totality, the universe respects the symmetry of the law that governs it; there is the same kind of behaviour on both sides of the threshold on which Janus traditionally stands. But should she, like Zeus's daughter Athena, come down from her Olympian vantage point to mix among mortals,[fn2] who can only exist on one side or other of the Janus point, she would find these ordinary beings perplexed. They cannot look through the Janus point and see what is on its other side; restricted in their vision, all they find around them are time-reversal-symmetric local laws of nature that, by some alchemy they call the past hypothesis, give rise to universally unidirectional arrows of time. But there is no alchemy. The law of the universe dictates the existence of the Janus point in all its solutions; there is symmetry overall, not in the particular.

BOTH WAYS IN which a Janus point can be realised, with or without the size of the universe becoming zero, require a radically new way of thinking about the nature of time and the universe. I introduced a little earlier a major theme in the book: that instead of regarding locally determined laws of nature as fundamental, we should instead recognise them as manifestations of a single, overarching law of the universe. There is justification for this. Ernst Mach, the great nineteenth-century physicist whose discovery of shock waves is honoured by the eponymous numbers measuring them, argued in a famous book on mechanics, published in 1883, that without historical investigation of science, “the principles treasured up in it become a system of half-understood prescripts, or worse, a system of prejudices. Historical investigation”, he said, “not only promotes the understanding of that which now is, but also brings new possibilities before us, by showing that which exists to be in great measure conventional and accidental. From the higher point of view at which different paths of thought converge we may look about us with freer vision and discover routes before unknown”. Mach’s book, with its critique of Newton’s notions of absolute space and time, was a major stimulus behind Einstein’s creation of the general theory of relativity between 1907 and 1915 as well as Heisenberg’s formulation of quantum mechanics in 1925.

So I take Mach’s admonition as encouragement for my investigation of the history of thermodynamics and the arrows of time. And the history matters. The purely practical and contingent circumstance of the birth of thermodynamics as a theory to explain the workings of steam engines has had a lasting effect: almost without exception, some means of confinement has, ever since Carnot’s pioneering study, played a critical role in the study and discussion of entropy and its associated arrow of time. In experimental work some real physical container, whose size can often be changed within finite limits (as when a piston is pushed into a cylinder), has been employed; in theoretical studies a box has been modelled as an essential part of the conceptual framework. However, results obtained within an enclosure are still widely and erroneously assumed to be valid in situations without an enclosure, even by scientists as distinguished as the late Richard Feynman. The only possible explanation for this is a lack of awareness of the critical condition—confinement and insulation—that must hold. If Feynman got it wrong, we might need to reconsider some things along the lines Mach advocated.

This does not mean blanket rejection of confinement. Far from it—an enclosure, both physical and conceptual, made possible some of the most profound discoveries in science, including unambiguous evidence for the existence and size of atoms and, in 1900, the discovery of quantum effects. The wonder and importance of that can hardly be overemphasised. But I will explain why, seventy years or so after Carnot published his book, the box scenario thwarted Boltzmann’s heroic attempts to give a completely convincing mechanical explanation of entropy’s increase. And I will argue that now, another 120 years on, a box mentality may still be blocking the way, through the Janus-point idea, to a true explanation not only of entropy growth but of all of time’s arrows.

If correct, this idea will resolve what appears to be a flagrant contradiction: the current consensus is that, for some as yet inexplicable reason, the universe began with very low entropy and a correspondingly high order that has since been undergoing inevitable progressive degradation through the inescapable effect of the second law of thermodynamics, yet all around us in the universe we see a wonderful growth of structure and order. One main aim of this book is to cast doubt on the widely accepted account of what has been happening throughout cosmic history, or at least on the notion of progressive degradation. I will use the example of records—for example, documents, fossils, and stars (considered as ‘fossils’ of stellar evolution)—which have remarkable mutual consistency, and give evidence for quite the opposite of what is claimed: that the history of the universe is not one of increasing disorder but rather of the growth of structure.

TO MAKE THIS case, let me begin with a personal example. In 1963 I was very lucky to buy clearly passed and are passing through a life cycle in which they develop an ever richer structure through the creation of atoms of ever increasing complexity—the stardust of which we are made. All this is now well understood on the basis of physical laws and observation. Astronomers' ability to understand these things now extends to the furthest reaches of the observable universe, both in space and in time. And all the evidence for this exists now, at this current cosmic epoch. In contrast to the stars, the development of galaxies is much less well understood. It's revealing that astronomers use stars as 'fossil records' and, seeking to understand the history of our galaxy, practise what they call galactic archaeology.

There is no need to multiply examples. We have learned many things about the world and the universe in which we live. One of the most remarkable is the cornucopia of records that, hanging together with great mutual consistency, tell a story of a universe of ever increasing complexity and structural richness. Appropriately analysed and interpreted, a grain of sand can now allow scientists to discern the history of the universe.

How do these stable records emerge, on both large and small scales, from the flux of history? How does it come about that the universe now is one huge time capsule of almost incomprehensible richness? I don't know the background to Muriel Rukeyser's motto in the epigraph to this chapter. Perhaps, like William Blake, she was reacting to the way science seems to drain joy and vitality from existence, replacing them with mere numbers. In fact, the billiard-ball-like atoms poets find so threatening are not at all like the atoms which quantum mechanics has discovered. In his *A Beautiful Question*, the physics Nobel laureate Frank Wilzcek holds that atoms are, literally, musical instruments.

By quoting Rukeyser, I don't in any way want to imply that atoms are the wrong basis for understanding the universe. We certainly cannot understand anything—our bodies, our brains, our phones, the aeroplanes we fly in—without the notion of atoms physicists now have. But what is the essential structure of any good story? It surely must have a beginning, an exposition, and an end. On that basis, the history of the universe is one huge story with countless diverse subplots all preserved in encoded form as records in the universe now. The universe is made of stories; scientists have learned how to decode and read them.

It turns out that they all unfold in a common direction. Bodley in Oxford is one of the world's great libraries, but we have been born into the ultimate library. It's all around us. It's the universe. One can argue, and I do, that the time-capsule nature of the universe— the way its state now proclaims a past history of growing complexity with extraordinary internal consistency—is the single most striking empirical fact that calls for scientific explanation. As of now, science can't provide it. All the known relevant laws of nature make no distinction whatsoever between the two possible directions of time. That's the rub. The laws tell us the timeline of the universe should be a two-lane highway. But all the stories I have mentioned—and many, many more—effectively erect a huge sign: THIS IS A ONE-WAY STREET.

I have presented this paean to the universe and its history not merely to emphasise the last point but rather as preparation for the claim this book makes. It is not just that it proposes a new theory of time and with it a first-principles explanation of its arrows; it also aims to overturn the doctrine that it is entropic disorder on a cosmic scale that puts the direction into time. The claim is this: the direction gets into time not through the growth of disorder but through the growth of structure and complexity. If that seems a rash claim, I invite you to go online and search for NASA's Timeline of the Universe. This and related images you will find at the same site show that at the big-bang birth of time there was a bland uniformity, from which there emerged, one by one, the most beautiful structures. In some you will see *Homo sapiens* as the current terminus; others show that terminus as the space probe WMAP, which a couple of decades ago made exquisitely accurate measurements of the very-low-frequency thermal radiation that bathes the universe and gives so many vital clues about our origin and place in the universe.

Suppose you show these images to the proverbial person in the street and say it's a story of the remorseless growth of disorder. I think the response might be 'Tell it to the marines'. If you consult a cosmologist interested in entropic questions, I think you will be told it's all something to do with the universe having a current entropy which is lower than the maximum it could have. This means the universe has an effective supply of free energy, which in thermodynamics as developed in the nineteenth century is the difference between the entropy that a system has at a given time and the maximum entropy it will have when no more useful work can be extracted from it. This is an account which is more or less compatible within conventional thermodynamics, though I think defining the entropy of an expanding universe as if the universe were a box of equilibrating gas is problematic to say the least. More seriously, I think that account shows only that entropy increase is not incompatible with the growth of structure. It does not show the direct connection between expansion of the universe and the way in which that expansion allows—indeed, mandates—the formation of specific structures. Above all, without the Janus-point idea it cannot explain the special initial condition that is needed, and therefore it cannot explain time's arrows.

CHAPTER 2

BARE-BONES THERMODYNAMICS

Heat and the First Law of Thermodynamics

In barely six pages, in which he proposed a method, albeit highly idealised, by which a steam engine could be made to operate with maximum efficiency, Carnot laid out all but one of the foundational principles of thermodynamics. His one mistake, which did not invalidate his proposal (known subsequently as the Carnot cycle), was to retain the contemporary theory of heat, which held it to be *caloric*, a weightless, incompressible, invisible fluid that could ‘insinuate’—Maxwell’s word—itself into bodies and cause them to expand. The mysterious substance came into physics in the late eighteenth century through a chance observation by Joseph Black, a professor at Edinburgh and adviser to whisky distillers. He happened to leave in a room two buckets; one contained ice and water, the other only water at its freezing point. A few hours later he found the second bucket’s water noticeably warmer and in the first less ice and correspondingly more water but, to his surprise, still at the freezing point. Only when all the ice had melted did that water begin to get warmer.

Ambient heat in the room had obviously brought about the changes. But why had it only melted some of the ice in the first bucket without heating the water as well? It seemed heat could disappear. Believing its amount should remain constant, Black suggested it had become latent (Latin for ‘hidden’). He also found that heating boiling water, while hastening evaporation, did not raise its temperature, which also suggested the heat had become hidden. To this day, students learn about latent heat, but now only as the heat needed to bring about phase transformations, such as the melting of ice or the boiling of water.

Carnot, for his part, likened the motive power of heat, presumed to be caloric, to that of water passing from a higher level to a lower one, driving a water mill in the process. In a steam engine, caloric would flow from the furnace at a high temperature into the cylinder, causing the steam in it to expand and move a piston. The cylinder would then be doused with cold water, which would take up caloric; with the caloric no longer present in the steam, the piston would return to the original position for the next cycle. Thus, according to Carnot, the motive power in steam engines is due “not to actual consumption of caloric, but to its transportation from a warm body to a cold body”. Just as a waterfall carries water from a height to a depth, the steam engine carries heat in the form of caloric from a high temperature to a low one.

This all too plausible picture had nevertheless been effectively killed a quarter of a century earlier by the colourful and resourceful American-born British physicist and inventor Sir Benjamin Thompson, Count von Rumford, while supervising the boring of cannons in Munich for the ruler of Bavaria. He showed that an effectively inexhaustible amount of heat could be generated through the friction accompanying the boring. Bystanders were astonished “to see how, without fire, such a quantity of cold water could be heated and even brought to boiling”.

The great question was therefore the one that “has so often occupied the natural philosophers: What is heat?” Having shown that the heat excited by the friction is inexhaustible and therefore could not possibly be a fixed amount of material substance, Rumford concluded: “It must therefore be *motion*”. He published his findings in the august *Philosophical Transactions* of the Royal Society of London and a chemistry journal in Germany. Despite this exposure, few scientists took any note of it for decades. The multitude of exciting experimental discoveries made with increasing frequency in the early decades of the nineteenth century is the probable reason.

IT WAS JAMES Joule who above all gave caloric the coup de grâce. He was a brewer in Manchester and, with every minute he could spare, an avid amateur scientist. He’s famous for the simplest of experiments. He allowed the fall of weights to turn paddles in closed water-filled tanks. The churning of the water heated it by friction. Joule’s brilliance was in measuring the tiny increase of the water temperature—perhaps the exigencies of brewing had something to do with that. By comparison, determining the work done by the turning paddles, measured by the total difference in height of the weights, was simplicity itself. This led Joule to the mechanical equivalent of heat: how much work must be done to generate a definite amount of heat. It is one of the most important numbers in science. In fact, what Rumford had measured in horsepower was, appropriately converted, reasonably close to Joule’s pounds per foot fallen.

Joule’s paper, published in 1843, attracted little attention until he spoke at the British Association in Oxford in 1848. The twenty-four-year-old William Thomson, already an expert in measuring very small temperature differences, was there standing at the back. He distrusted Joule’s claim but, sufficiently intrigued, talked to Joule and corresponded with him. The result was prominent mention of the paddle experiment in Thomson’s 1849 paper, the one that put Carnot’s work on the map. Unfortunately for Thomson, he failed to register the critical importance of an earlier electrical experiment in which Joule had shown that heat could be transformed into mechanical work. Heat and work were completely interchangeable. Thomson later admitted somewhat ruefully in a letter to Joule his failure to appreciate that heat could not only be created but also be ‘put out of existence’ (by doing work). Failing to shake off Carnot’s waterfall analogy, Thomson opined in his influential 1849 paper that the notion of caloric should not be abandoned before several more years of careful measurement.

CLAUSIUS HAD NO such inhibition. In his paper of 1850, he argued persuasively that a steam engine is not a water mill. The heat that comes from the furnace is not conserved; some is converted into work. It does not become hidden—latent—but ceases to exist; as in Joule’s electrical experiment, nature has balanced things in the work done. The residual heat that remains in the working medium passes into the water sprayed onto the steam engine cylinder. Clausius’s modification, and with it the completion of Carnot’s theory, was the first clear example in which what is arguably the most fundamental physical law—the conservation of energy—is manifested. Thomson had missed that boat but soon still made an important contribution to the definitive formulation of the first law of thermodynamics by including in it the change in the internal energy of bodies. He was, in fact, the first person to establish widespread use of the word ‘energy’ and its two aspects: potential, as in the weight lifted or an apple on a tree, and kinetic, as the apple falls. The latter replaced the old name *vis viva* (living force). The difference between kinetic and potential energy in gravitational theory is going to be very important later in the book; it will be good to illustrate it by the example of a pendulum. When the bob is at its lowest point, its speed is greatest and the kinetic energy, which is always positive and equal to $mv^2/2$, where m is the mass of the bob and v is its speed, has its maximum value. At the highest point of its swing, the bob is at rest and has no kinetic energy. The bob’s potential energy, for its part, is always negative, becoming less so with increasing height in such a way that throughout the motion the sum of the kinetic and potential energies remains exactly constant. As the pendulum swings back and forth, there is therefore a constant ebb and flow between the two forms of energy, subject to strict constancy of the total.

The discovery and formulation of the first law made a huge impression in what was still a profoundly religious age. The idea that God had created a universe in which, despite the vagaries of existence, something remains eternally unchanged was very comforting. Joule, as religious as Thomson, had already expressed the sentiment in 1847. As reported in the *Manchester Courier*, a talk he gave in that year brimmed with the confidence of a man who knows he had made a great discovery. Joule referred to one of the most important qualities with which God had endowed matter and that it would be absurd to suppose it could be destroyed or created by human agency.

Inviting his audience to behold “the wonderful arrangements of creation” and the vast variety of phenomena involving the conversion of living force and heat into one another, he said that they “speak in language which cannot be misunderstood of the wisdom and beneficence of the Great Architect of nature”. It is something we see “in our own animal frames, ‘fearfully and wonderfully made’”. Indeed,

> the phenomena of nature, whether mechanical, chemical, or vital, consist almost entirely in a continual conversion of attraction through space, living force, and heat into one another. Thus it is that order is maintained in the universe—nothing is deranged, nothing ever lost, but the entire machinery, complicated as it is, works smoothly and harmoniously ... the whole being governed by the sovereign will of God.

Coming back down to earth, it’s a nice thought that caloric came into science through an adviser to distillers of whisky and went out through the efforts of a brewer of beer.

The Second Law, Dissipation of Energy, and Entropy

One of Carnot’s inspirations was that heat was universal in its effect: whatever the working medium might be—steam or some simple gas—the efficiency of a steam engine operated under ideal circumstances would depend solely on the amount and temperature of the heat taken from the furnace and the amount and temperature at which it was cooled by dousing. He argued, by impeccable logic, that if this were not the case it would be possible to build a perpetual-motion machine from which one could extract an unlimited amount of work. But the claims of charlatans that they had functions; they characterise the equilibrium state. Any two fix the third and with it the equilibrium state. Temperature is a particularly interesting state function. Since the early eighteenth century the ideal-gas law had suggested something special must happen at about –273°C. This became known as the absolute zero of temperature through the theoretical work of Thomson, who, building on Carnot's insights, showed that the mechanical effect of heat could be used to measure temperature. This eliminated the problem that the thermometers hitherto used, which were based on expansion of gases and liquids, did not give concordant results. Thomson's measurements also determined the absolute zero of temperature with great accuracy, in recognition of which absolute temperatures are measured in kelvins.

IF CALORIC DID exist and could be neither created nor destroyed, it too would, like pressure and temperature, be a state function. But Joule's experiments had done away with caloric. Already in 1850 Clausius began to suspect some other state function might take its place. By 1854 he had it.

Carnot's idealised steam engine required heat to be transferred from a furnace to steam in a cylinder at a temperature only infinitesimally lower. The steam would be allowed to expand very gradually, thereby raising infinitesimal weights that could be removed one by one at successive heights. In the ideal limit, this process is reversible: one by one the weights are put back on the piston, causing it to sink and raise the temperature of the steam ever so slightly above that of the furnace, into which heat would now flow. Were heat to be transferred from the furnace with finite temperature difference, some of the heat would simply raise the temperature of the steam without being fully exploited to do work. As Thomson had noted in 1852, Carnot placed great stress on the need to ensure that the maximum possible amount of work should be done whenever heat passed from a higher temperature to a lower one. Water mills worked on the same principle—no water should fall without doing some work.

The identification of reversible transformations was one of Carnot's several great contributions to science. Clausius took his work further into a quite new realm. His great idea was that one could gently 'nudge' enclosed gas from one equilibrium state to another by letting it exchange infinitesimal amounts of heat with an adjacent heat reservoir at an infinitesimally higher or lower temperature. This insight came straight from Carnot.

Here I must introduce two or three of the symbols that Clausius employed; they have become standard. By Q he denoted an amount of heat and by $\mathrm{d}Q$ an infinitesimal amount of it. A reversible transformation then involves either the infinitesimal transfer of $\mathrm{d}Q$ of heat to the gas or $-\mathrm{d}Q$ from it. The decisive step was that Clausius then divided the amount of heat $\mathrm{d}Q$ by the absolute temperature T and called the resulting expression

the (infinitesimal) transformation value of the process. It was only in 1865 that, with aplomb, Clausius replaced the clumsy phrase 'transformation value' by 'entropy', which he still denoted by S and the infinitesimal increment of entropy by $\mathrm{d}S$. It's possible that, to honour Carnot, he chose S for Carnot's first name, Sadi. It's a nice idea.

Your next question might be to know the significance of the division of the amount of heat $\mathrm{d}Q$ by the absolute temperature T. Well, Joule's experiments had shown that all heat can do work, so $\mathrm{d}Q$ could, say, lift eight ounces through ten inches. Now temperature measures something like the ‘quality’ of heat. You feel that when coming from the cold into a warm room. In fact, if you are feeling cold and have the option of going into one of two rooms, one large and one small and both containing the same amount of heat, you will surely go into the small and hotter room.

Reflecting that, the increment of entropy dS in Clausius’s definition measures the quality of the added heat dQ. If it is added at a low temperature T, then the entropy increment dS = dQ/T is large; the heat is, as it were, ‘spread out’ or diffuse, as indeed it is in the large room you would not choose to enter. If T is large, dS is small, the heat is more ‘concentrated’ and it has higher quality. You can achieve more with it.

I MUST NOT omit to illustrate here this matter of the quality of heat by the most important engineering insight that emerged through the work of Carnot, Thomson, and Clausius. The Frenchman had proved that the maximum possible efficiency of a steam engine depends solely on the two temperatures between which it operates: the temperature of the steam heated by the furnace, T_{steam}, and the temperature of the steam-dousing coolant, T_{coolant}. However, he had not been able to find an expression for the maximum possible efficiency. Mainly through Thomson’s definition of the absolute scale of temperature, it was shown to be the difference between these two temperatures divided by the absolute temperature T_{steam} of the steam. This then is the great triumph of thermodynamics: the maximum possible efficiency of a steam engine, typically denoted by the Greek letter η (eta), is

Unless the steam is superheated, its temperature is 373 kelvin, while the coolant will typically be at the ambient temperature, say 293 kelvin. This makes the maximum efficiency

which is only a bit over 21 per cent. Most steam engines work with a much lower efficiency. The very first ones achieved barely 2 per cent. You might like to go online to read about the first commercially successful steam engine invented by Thomas Newcomen in 1711 to pump water from mines. Note that the formula for the efficiency expresses it as a ratio. This exemplifies the fact, noted in Chapter 1, that it is only ratios that have physical significance.

SO FAR I have only told you how Clausius defined the increment of entropy; I have not described how he defined entropy itself or what he considered to be its full significance. Let’s start with the definition. Clausius supposed some closed system, for example ideal gas in a cylinder whose volume can be changed by a piston. The gas is initially in some equilibrium state, to which Clausius ascribed a nominal entropy S_{initial}. He then imagined the gas taken by infinitesimal reversible transformations through a succession of equilibrium states, each with its own pressure, temperature, and volume. The only thing he took into account was the heat taken from or given up to a series of heat reservoirs by the gas at each stage of his delicate manoeuvre. Any work the gas might do through the piston or the piston might do to it did not enter his calculations. He was solely concerned with entropy increments dS. They should be added to (for a positive amount of heat, + d*Q*) or subtracted from (for a negative amount, –d*Q*) the current value. Then after *N* steps he arrived at a final value of the entropy:

Clausius proved that no matter which of the infinitely many ‘paths’ through the possible equilibrium states was chosen, S_{final} would always be the same and depend only on the initial and final equilibrium states. If they, along with the nominal value of S_{initial}, are specified, then the difference $S_{\text{final}} - S_{\text{initial}}$, and with it S_{final}, is fixed. It is path-independent.

The notion of path independence is one of the most important in both mathematics, the queen of the sciences, and physics. I therefore owe you at least some kind of explanation of what it means. Luckily it’s not too difficult. Consider first the surface of the earth. At every location in a given country, the height above mean sea level, the altitude, has a fixed value. If you go by any path from one given location to another, your altitude will change by the same amount whatever route you take. The difference in altitude is path independent. But now suppose the country has toll roads with variable rates, and on some you make payments but on others you are given bonus payments to encourage travel along routes which do less harm to the environment. Then obviously the cost of a journey will depend on the route you take. It will be path dependent. I think you will agree that the basic idea is simple even if its applications can be very subtle.

It was in Clausius’s case. His proof that the difference between S_{final} and S_{initial} depends only on the initial and final equilibrium states and not on the path taken between them is extraordinarily clever and involves heat transfer not only to or from the *N* heat reservoirs but also to and from just as many heat engines. Examples of these are practical things like refrigerators, the first ‘thermodynamic spinoffs’, brilliantly exploited by Thomson wearing his engineer’s hat in the Joule-Kelvin process, as well as Carnot’s model of an idealised steam engine working at maximum efficiency. Besides all this ‘auxiliary machinery’, Clausius employed one further heat source at a uniform fixed temperature. It is there so that, at the critical point in the argument, it can be argued that if S_{final} were not path independent, work would have been done in violation of the second law: by extracting heat from a uniform source (Thomson’s form of the second law as used by Fermi) or, as in Clausius’s original proof, heat would have flowed spontaneously from a colder body to a hotter one. The power of the second law in either of these utterly simple formulations is breathtaking. Fermi uses it twice to establish the absolute thermodynamic scale of temperature and once more to prove the path independence of S_{final}.

This is what enabled Clausius to argue that S_{final}, not yet called entropy, is a state function. Its dependence on S_{initial}, which can be fixed once and for all by choosing a standard reference state, is of no concern. The real thing of interest was the change, the resulting difference, in going from one state to another. What Clausius had found is a quantity that characterises the capacity of a considered substance to do useful work. The temperature tells you what it feels like; the new state function tells you something about what you can do with it. I should also say that although the nominal value given to S_{initial} is immaterial, some reference state and value must be chosen.