OXFORD

ENERGY

the Subtle Concept

The discovery of Feynman's blocks from Leibniz to Einstein

JENNIFER COOPERSMITH

ENERGY, THE SUBTLE CONCEPT

Nicolas Poussin, 'A Dance to the Music of Time', *c*. 1640, by permission of the Trustees of the Wallace Collection, London. A metaphor for the ever-changing forms of energy.

Energy, the Subtle Concept

The discovery of Feynman's blocks from Leibniz to Einstein

JENNIFER COOPERSMITH

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2010

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire

ISBN 978-0-19-954650-3

1 3 5 7 9 10 8 6 4 2

To Bertie Coopersmith and Murray Peake, sages both

This page intentionally left blank

Contents

Preface

The aim of this book is to explain energy and to use the history of its emergence for this purpose. As the book is therefore primarily physics-explanation rather than history of physics, no attempt is made at a comprehensive historical account. The chronology is broken up by frequent asides and explanations in modern terms. The amount of biographical detail is mostly in inverse proportion to the fame of the given scientist and many important scientists are left out altogether (Colding, Rankine, Séguin, Lomonosov, la Borda, Bélidor, Holtzmann, Krönig, Ewart, Hachette, Coriolis, and others).

Chapters 2–16 chart the discovery of energy up to the mid-nineteenth century. Chapters 17 and 18 look forward to energy in the modern era: the treatment is more condensed and assumes a background in the physical sciences.

In order to avoid a forest of quotation marks, I have started off with them (as in ‘energy’, ‘force’, ‘physicist’) and then dropped them quickly, with little attempt at consistency.

The symbol ‘$\propto$’ means ‘proportional to’, ‘Δx’ means ‘a small increment in x’, and ‘<*x*>’ means ‘the average value of x’. Only as the twentieth century is reached do I indicate vector quantities (by writing them in bold).

Acknowledgements

Six marvellous books have guided me and been my constant companions: Donald Cardwell's *From Watt to Clausius: the Rise of Thermodynamics in the Early Industrial Age*, Cornelius Lanczos' *The Variational Principles of Mechanics*, Richard Feynman's *Lectures on Physics*, Richard Westfall's *Force in Newton's Physics*, Charles Gillispie's *The Edge of Objectivity*, and Brian Pippard's *The Elements of Classical Thermodynamics*.

In the case of the last two books, I have been privileged to have a correspondence with the authors. Both have been extremely enthusiastic and encouraging (Professor Sir Brian Pippard died last year). Professor Gillispie (Emeritus Professor at Princeton and founding father of the discipline of the history of science) has read many of my draft chapters and made invaluable suggestions. Professor Paul Davies (physicist, science writer, and founder and director of 'Beyond, Center for Fundamental Concepts in Science'), and Emeritus Professor Rod Home (Department of History and Philosophy of Science, Melbourne University) read some early draft chapters and were likewise very encouraging.

I would like to thank past colleagues and friends at King's College London, TRIUMF, UBC, Logica SDS, and the Open University. I would especially like to thank the Open University students (London, Winchester, and Oxford, 1986–96) who asked all those difficult-to-answer questions.

My present colleagues at La Trobe University, Bendigo Campus, and Swinburne University of Technology in Melbourne have been very supportive. In particular, I would like to thank Rob Glaisher, Katherine Legge, Glenys Shirley, John Schutz, Mal Haysom, Joe Petrolito, John Russell, Glen Mackie, and Sarah Maddison. For special tasks: Mal Haysom took the cover photos, Andrew Kilpatrick was tireless in drawing diagrams, even randomizing the position of dot-molecules by hand (mind you, I assisted with his marking-avoidance strategies), Sabine Wilkens translated Euler and trawled through Helmholtz's endless accounts of indigestion, and Glenys Shirley helped with anything and everything, and with characteristic good humour.

I thank the University of La Trobe, Faculty of Science, Technology and Engineering, Bendigo Campus, for providing me with an office, computing and library facilities, and the intangible but crucially important ambience of learning, research, and cooperation.

The library staff at La Trobe University, Bendigo, showed great forbearance in my constant requests for inter-library loans and my loss of magnetic belts, late returns, and so on. (Having previously lived a stone's throw from the Blackwell bookshop and the Bodleian in Oxford, I naturally didn't embark on this project until I had moved to a part of Victoria in which there are more kangaroos than bookshops or libraries.)

I thank Howard Doherty, of Doherty's Garage in Bendigo, for interesting conversations about cars, their design, and their impact on society.

Thank you to Wikimedia Commons and Google Books for making source documents available online. In material reproduced in this book, every reasonable effort has been made to contact the original copyright holders. Any omissions will be rectified in future printings if notice is given to the publisher.

I am unusually fortunate in having not one but two sages to call upon in my family (I am daughter of one and married to the other).

My father (a physicist-turned-programmer) is a deep and highly original thinker, always aiming to get back to the essentials. When I asked him where my grandfather had come from in Russia and what the family name was (before 'Coopersmith' was adopted upon arrival in South Africa around 1910), he replied 'I don't know. Anyway it's just contingent.'

My husband, Murray, has a profound understanding of mathematics and physics (he was a prize-winning physics student and applied mathematician before being drawn down that well-trodden path of computer programming). His compendious knowledge of pre-twentieth-century arts and crafts has also been useful in the writing of this book. After one of his weekday 'Murraycles' (for example, being instrumental in a quiet revolution in the calculation of risk in banking), he turns to tying knots, grinding wheat, making rope, bread, tofu, and soap, sewing, and restoring gas lamps and kerosene burners. He is an unassuming polymath, so why switch on the computer or find the calculator when you can always ask 'Murraypaedia' or get him to do a back-of-the-envelope calculation with the appropriate series approximation? His large collection of books ranges from *Spruce-root Basketry in the Tinglit* to *Galactic Dynamics*, and he has a knack for swiftly finding the very book needed to solve one's problem. Together, we usually manage to complete Araucaria. The fact that he is also a teddy bear of a man with gentle, blue eyes is just the icing on the cake.

My mother has frequently held the fort (or fortress, as our Dutch uncle calls it) with inimitable Granny-ish style and vigour. She is nevertheless impervious to my constant exhortations: 'You don't need to practice worrying—when you need to worry you'll know how.'

I would like to thank the copy-editor, Geoffrey Palmer, whose professionalism and background in science were invaluable. Also Jan Wijninckx (the Dutch uncle) was very generous with his time and software expertise and saved me from some Word-wracking moments.

Many thanks also to the 'patchwork quilt' of childcare, most especially Helen and Richard Jordan and Clare and Phil Robertson.

Finally, I thank my children, David, Rachel, and Deborah, for helping in all the usual ways; in other words, without their efforts this book could have been completed in half the time. Seriously, though, they have been very appreciative of this enterprise and provided a wonderful selection of live and recorded music to write books to.

List of Illustrations

1

Introduction: Feynman's Blocks

Ask a physicist what physics is all about and he or she might reply that it's something to do with the study of matter and energy. 'Matter' is dispensed with quite swiftly—it is stuff, substance, what things are made from. But 'energy' is a much more difficult idea. The physicist may mumble something about energy having many different forms, and about how it is convertible from one form to another but its total value must always remain constant. She'll look grave and say that this law, the law of the conservation of energy, is a very important law—perhaps the most important law in all of physics.

But what, exactly, is being conserved? Are some forms of energy more fundamental than others? What is the link between energy, space, time, and matter? The various forms and their formulae are so seemingly unrelated—is there some underlying essence of energy?

The aim of this book is to answer these questions and to *explain* the concept of energy *through the history* of its discovery. The validity of employing peculiarly anthropocentric concepts ('work', 'machine', 'engine', and 'efficiency') then becomes clear. More than this, the history shows that great philosophical and cultural—*intellectual*—revolutions were required and these were every bit as paradoxical, as discomforting, and as profound as those that occurred with the arrival of quantum mechanics and Einstein's Relativity. The long list of revolutions, *before* the twentieth century, includes the following: the physical world must be looked at and *measured*; mathematics is the language of physics; time is a physical parameter and can be compared with distance; the void exists; perpetual motion—of a continually acting machine—must be vetoed; perpetual motion—of an isolated body—must be sanctioned; the latter must be vetoed, *in practice*; nature must conserve her resources; nature must be economical with her resources; action-at-a-distance is absurd but true; the world is cyclical and deterministic; the world is not cyclical but progressive; there are two profoundly different kinds of motion—bulk and microscopic; subtle, weightless 'ethers' are essential; most subtle, weightless ethers are redundant; light is particulate; light is wave-like and has *transverse* components; not all forces are central; Newton's Third Law isn't always true; physics has statistical and probabilistic aspects; some laws are not absolutely true; light and matter can be treated by the same mechanics; electricity, magnetism, and light are all manifestations of the same thing; the electromagnetic field is *real*; and what's important is the system, the whole system, and nothing but the system (to roughly paraphrase d'Alembert's Principle).

Also, the historical sequence of events is surprising. The steam age was well under way—steam locomotives pulling trains and steam engines powering industry—decades before 'energy' had been discovered and its conservation stated as the First Law of Thermodynamics. Also, the Second Law trumped the First Law by being discovered first, then came the Third Law, and the Zeroth Law was last of all.

Finally, the historical approach is hugely entertaining. In fact, the history of scientific ideas—of knowledge—is the ultimate 'human interest' story; and it is a ceaseless wonder that our universal, *objective* science comes out of human—sometimes all too human—enquiry.

This is a tale of persecuted genius and of royal patronage, of ivory-towered professors and lowly laboratory assistants, of social climbers and other-worldly dreamers, of the richest man in all Ireland, the richest man in all England, and one who had to make a loaf of bread last a week. It includes feuding families and prodigal sons, the Thomson and Thompson 'twins', a foundling, a Lord, revolutionary aristocrats, entrepreneurs, and industrialists, clerics, lawyers, academics, engineers, *savants*, doctors, pharmacists, diplomats, a soldier, a teacher, a spy, a taxman, and a brewer. Some were lauded and became grandees of science, others only received recognition posthumously. A few were persecuted (under house arrest, imprisoned, one even guillotined, another lampooned), while others carried on through trying domestic circumstances (for example, Euler, when old and blind—he did the calculations in his head—was surrounded by a horde of grandchildren, all possessions lost in a house-fire, and yet maintained his average work-rate of one paper a week for 60 years). There was a ladies' man, a pathologically shy man, a gregarious American, and a very quiet American. Yes, they were all men, although two wives and a mistress put in an appearance (one wife actually appears twice, married first to Lavoisier and then to Count Rumford).

There were English eccentrics and gentlemen-scientists; Scottish 'Humeans', and, later, Scottish Presbyterians; French *philosophes* and, post-Revolution, French professional scientists; German Romantics, then materialists, and, later, positivists. Some were great travellers and others travelled only in their minds (Newton never even visited Oxford, let alone other countries). However, the community of scientists was small enough that all could know each other. Communication was obviously slow (for example, by stagecoach in the seventeenth and eighteenth centuries) yet, apart from one glaring exception, there was always a spirit of co-operation between scientists of different nationalities, even when wars were raging between the respective countries.

Some were what we would now call philosophers (Descartes and Leibniz), applied mathematicians (Euler), mathematicians (Lagrange and Hamilton), chemists (Boyle, Lavoisier, Scheele, Black, and Davy), and engineers (Watt and Carnot), and some were true physicists (Newton, Daniel Bernoulli, and Joule). Some were lovelorn and poetical (Hamilton and Davy), or dour and preoccupied with earning a living (Watt), and some were universal geniuses (Leibniz and Young). All were very hard-working.

There seems to be a conspiracy of silence surrounding topics such as the following: What is action? Where does the kinetic energy go to when we change frames of reference? Why the two forms, kinetic and potential energy? Why is $(T - V)$ significant and why must it be minimized? What is d'Alembert's Principle and why is it important? All relate to that most important principle—perhaps *the* most important principle in the whole of physics—that physics must be the same for all observers. These ideas came through in the century from Descartes to the French Revolution, the 'only period of cosmic thinking in the entire history of Europe since the time of the Greeks'[1] until resumed again with the work of Einstein in the twentieth century.

Other rarely explained topics are as follows: the meaning of entropy and why the famous (Boltzmann's) microscopic formula is *it*; why temperature is *the* parameter in thermodynamics and why it is universal; and the paradox of time's arrow (or, why things happen). All will be explained, as well as can be outside of a textbook or a laboratory and without much mathematics. Simply imagine that we're in conversation during a very long train journey or over endless cups of coffee in the common room.

Just how slippery the concept of energy is has been captured by Richard Feynman in an allegory in his *Lectures on Physics*.[2] (If Feynman says the concept of energy is difficult then you know it really *is* difficult.)

Feynman imagines a child, 'Dennis the Menace', playing with blocks. Dennis' mother discovers a rule—that there is always the same number of blocks, say 42, after every play-session. However, one day she finds only 41 (the last one was under the rug) and on another day she finds only 40 (two had been thrown out of the window). On yet another occasion, there were 45 (it turned out that a friend, Bruce, had left three of *his* blocks behind).

Banning Bruce and closing the window, everything is going well until one day there are only 39 blocks. The mother, being ingenious, finds that the locked toy box, the empty weight of which is 16 ounces, now weighs more. She also knows that a block weighs 3 ounces and so she discovers the following rule:

$$(\text{number of blocks seen}) + ([(\text{weight of box}) - 16 \text{ ounces}]/3 \text{ ounces}) = \text{constant}$$

On another occasion, the water in the bath has risen to a new level but is so dirty (!) that she can't see through it. She finds a new term that she must add to the left-hand side of the above equation: it is [(height of water) – 6 inches]/1/4 inch.

As Dennis finds more and more places in which to hide the blocks, so his mother's formula becomes more and more complicated, with a longer and longer list of seemingly unrelated terms. The complex formula is a quantity *that has to be computed* but that always adds to the same final value. Feynman concludes:

> What is the analogy of this to the conservation of energy? The most remarkable aspect that must be abstracted from this picture is that *there are no blocks*.[3]

Something is conserved—but what? There are two other lessons to be learned from Feynman's allegory: a system must be defined and kept isolated; and quantification and mathematization are essential. We shall attempt to understand the nature of these non-existent 'blocks' of energy through the history of their discovery.

2

Perpetual Motion

Energy was always energy even before it was understood as such—a force for making things happen, for driving every kind of process, device, or machine, a source of power for windmills and water-wheels. So, it will be profitable in our search for the 'blocks' of energy to look at the quantitative understanding of machines through history.[1]

The first attempt at a quantitative analysis was the description of the lever by the followers of Aristotle. These were the Peripatetics, so-called because they wandered around the gardens of the Lyceum in Athens while carrying on with their discussions. In the *Mechanica*, a work written by the Peripatetic school around 300 BC, the following question was posed:

> Why is it that small forces can move great weights by means of a lever?[2]

The answer followed:

> the ratio of the weight moved to the weight moving it is the inverse ratio of the distances from the centre the...point further from the centre describes the greater circle, so that by use of the same force, when the motive force is further from the lever, it will cause a greater movement.

This sounds quite modern until we examine the reasons given in more detail:

> The original cause of all such phenomena is the circle. It is quite natural that this should be so; for there is nothing strange in a lesser marvel being caused by a greater marvel, and it is a very great marvel that contraries should be present together, and the circle is made up of contraries.

The contraries were 'motion and rest' (the line of the circle moves, while the point at the centre is at rest); 'the concave and the convex' (the line of the circle defines both a concave and a convex surface); and one circle turning in one direction causes an adjacent circle to rotate in the opposite direction.

The most famous proof of the law of the lever is that due to Archimedes (around 250 BC in Syracuse). He found the constraint on the location of the centre of gravity on either side of the fulcrum when the lever was in equilibrium. (He takes 'centre of gravity' as a self-evident idea, needing no further justification.) Legend has it (or, rather, Pappus of Alexandria, *c.* AD 340) that Archimedes said 'Give me a place to stand on, and I will move the Earth.'

The enigmatic Hero of Alexandria lived around AD 60, and devised many ingenious machines and 'toys' for use in temples and as spectacles for an audience. He was perhaps the first to use steam power—for a toy, a sort of kettle with two bent spouts and supported on mounts that allowed the kettle to keep on turning as the steam escaped.

(The Chinese used steam in antiquity, but in laundry presses rather than as a source of power.[3]) Hero listed the five simple machines as the wheel and axle, the lever, the pulley, the wedge, and the inclined plane. He understood that all were variations on one elemental machine and that all obeyed the fundamental law: 'Force is to force as time is to time, inversely.'[4]

Not much happened (as regards the mathematization of machines) for well over 1,000 years. Around AD 1250, Jordanus Nemorarius derived the law of the lever again and was the first to use the concept of virtual displacements (to be re-invented by Johann Bernoulli in the eighteenth century; see Chapter 7, Part IV).

The semi-legendary Chinese figure Ko Yu is credited with the invention of the wheelbarrow around AD 1. Later, these were often sail-powered (a wheelbarrow with a sail was a popular image for China in eighteenth-century Europe). A window in Chartres Cathedral, dated 1225, has the earliest depiction of a wheelbarrow in use in the West.

The first water-wheel was found in China around 200 BC. The development of gears at about this time enabled the wheel to be ox-powered and used for irrigation. In Illyria (present-day Albania) in around 100 BC, water-powered mills were used for grinding corn. By 1086, in England, the Domesday book listed 5,624 water-wheel mills south of the River Trent, or one for every 400 people.

The earliest known windmills were used in Persia (Iran) in around AD 600 (AD will be assumed from now on). They had a vertical shaft and horizontal sails and were used to grind grain. Windmills—for draining the land—were found in Holland from around 1400.

But what happened when the rivers dried up and the wind didn't blow? Since ancient times, there have been continuing attempts to make a machine that, once started, would run forever—a perpetual motion machine (or *perpetuum mobile* in Latin, the language of culture in Europe since the Roman Empire).

The idea of a perpetually acting device appears to have originated in India, where a perpetually rotating wheel had religious significance, symbolizing eternal cycles such as the cycle of reincarnation (wheel symbols often appear in Indian temples).

The first mention of a perpetual motion machine occurs in the Brahmasphutassiddhanta, a Sanskrit text by the Indian mathematician and astronomer Brahmagupta in 624. He writes: 'Make a wheel of light timber, with uniformly hollow spokes at equal intervals. Fill each spoke up to half with mercury and seal its opening situated in the rim. Set up the wheel so that its axle rests horizontally on two supports. Then the mercury runs upwards [in some] hollow spaces and downwards [in some others, as a result of which] the wheel rotates automatically forever.'[5] Subsequent texts by the Indian astronomers Lalla in 748 and Bhaskara II in 1150 describe similar wheels. (Bhaskara II is also famous as the first mathematician to define division by zero as leading to infinity.)

Such over-balancing wheels were to prove one of the most popular methods in the attempt to achieve perpetual motion. The French architect Villard (Wilars) de Honecourt tried it in 1235:

> Now there are major disputes about how to make a wheel turn of itself: here is a way one could do this by means of an uneven number of mallets or by quicksilver.[6] (See Fig. 2.1 overpage)

Edward Somerset (*c.* 1601–67), the Marquis of Worcester, famously made an over-balancing wheel while held in the Tower of London after the beheading of Charles I. The wheel was made on a huge scale ('fourteen foot over, and forty weights of fifty pounds apiece'[7]) and probably kept going for some considerable time by acting as a flywheel. King Charles II was sufficiently impressed to release the Marquis, who subsequently developed his 'water-commanding engine', the very first steam engine (Chapter 4).

As remarked earlier, sometimes the rivers dried up and sometimes the wind didn't blow. Inventors then wondered if a fixed supply of water or air could drive a machine in a closed-cycle operation. Robert Fludd (1574–1637), physician and mystic, proposed a closed-cycle water-mill in 1618.

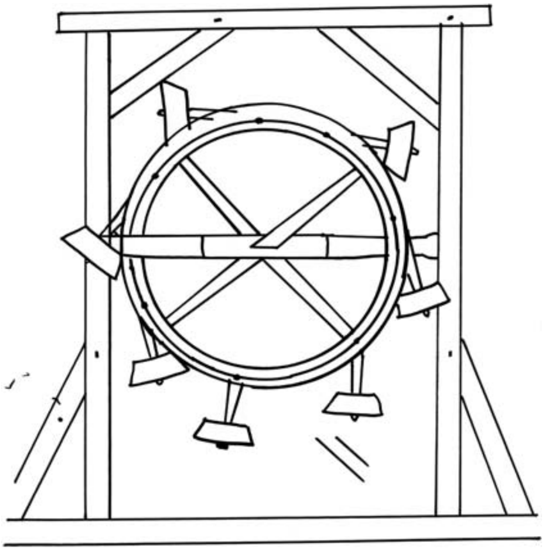

Fig. 2.1 Villard de Honecourt's over-balancing wheel, 1235 (courtesy of Wikimedia Commons).

More obviously problematic was Mark Antony Zimara's self-blowing windmill from 100 years earlier (1518).[8] The impression by the artist Lee Potterveld, commissioned by Professor Kasten Tallmadge (Fig. 2.2), is in a style appropriate to the age. Could the vanes of the windmill ever hope to operate those huge bellows?

John Wilkins (1614–72), Bishop of Chester and an early official of the Royal Society (founded in London in 1660), proposed three sources that might lead to perpetual motion: 'Chymical Extractions', 'Magnetical Virtues' (Fig. 2.3), and 'the Natural Affection of Gravity'.[9] Wilkins was part of the seventeenth-century scientific establishment—clearly such ideas were not considered far-fetched.

Another proposed source of perpetual motion was capillary action, as in Boyle's and also Papin's 'goblet' in the seventeenth century (Fig. 2.4). (Robert Boyle was the famous 'chemyst' and discoverer of Boyle's Law; Denis Papin was famous for his pressure cooker or 'digester'—see Chapter 4.)

Vittorio Zonca's perpetually siphoning closed-cycle mill of 1656 (Fig. 2.5) was a similar idea.

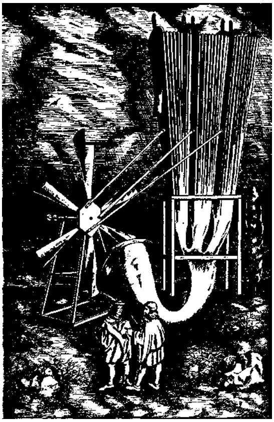

Fig. 2.2 Artist Burton Lee Potterveld's impression of Zimara's self-blowing windmill, 1518 (with the permission of Lee Potterveld's son).

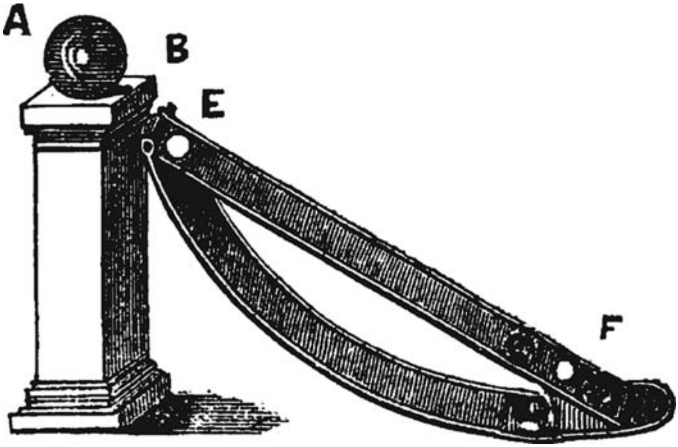

Fig. 2.3 Bishop Wilkins' magnetic perpetual motion, 1648, from *Perpetuum Mobile, or the Search for Self-motive Power during the Seventeenth, Eighteenth and Nineteenth Centuries*, by Henry Dircks (1861).

Johann Bernoulli (1667–1748), of the famous Bernoulli family of mathematicians (see Chapter 7), proposed perpetual motion using a sophisticated but flawed hydrostatic analysis of a system of two fluids. Sir William Congreve (1772–1828), politician and inventor of the

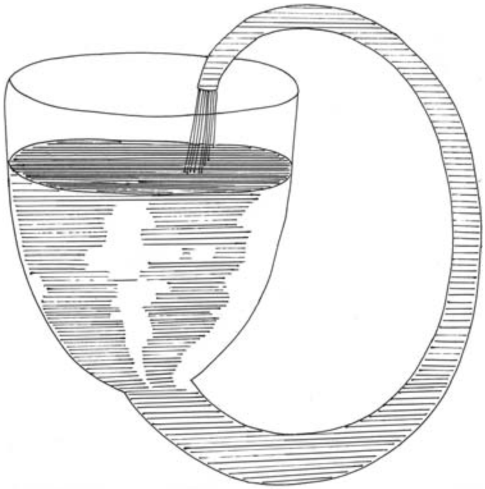

Fig. 2.4 Boyle's and Papin's 'goblet' (after Dircks 1861, as in Fig. 2.3).

Congreve rocket, was amongst a number of perpetual motionists in the eighteenth century who utilized sponges and osmosis; and there were others who employed capillary tubes within capillaries within capillaries, and so on.

By the eighteenth century, more and more charlatans and fraudsters were joining the ranks of genuine attempts at perpetual motion. In the tale of Orffyreus—a pseudonym for the German entrepreneur Johann Bessler (1680–1745)—a perpetually moving wheel was set up for display in Hesse Castle, Kassel. However, the interior mechanism of the wheel was concealed and the wheel itself kept in a locked room, with the Landgrave of Hesse's seal upon the door. The Landgrave allowed the Dutch natural philosopher Willem 's Gravesande of Leyden (1688–1742; see Chapter 7) to inspect the wheel, but without opening up the mechanism. Professor 's Gravesande subsequently sent a report of the wheel to Isaac Newton. It is not known whether or not Newton was impressed, but Orffyreus was apparently so outraged that he smashed the wheel to bits and scrawled a message on the wall saying that he was driven to wreck his wheel by 's Gravesande's impertinence.

Newton justified his Third Law of Motion (Chapter 3) by the evident absurdity of perpetual motion, yet, in his youthful entries in his

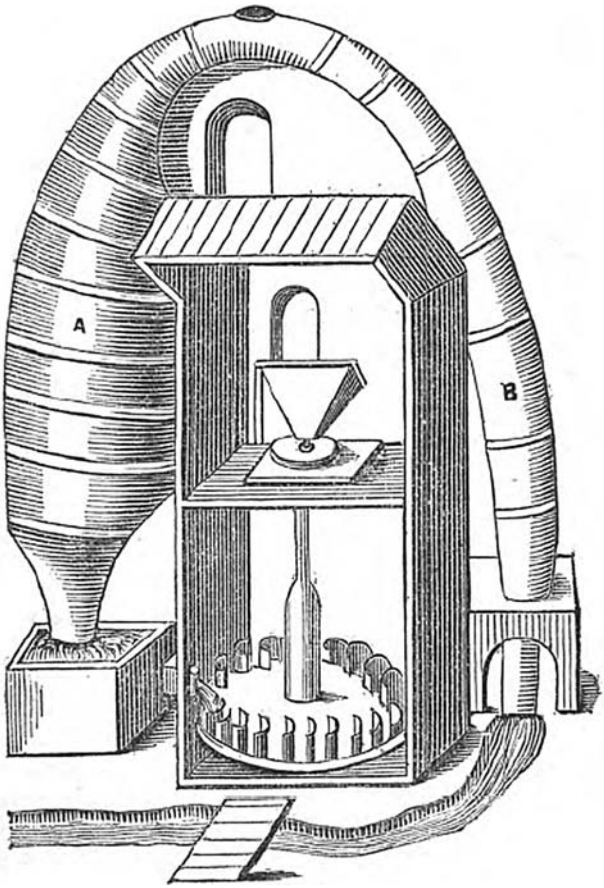

Fig. 2.5 Zonca's siphon, in *Nuovo Teatro di Machine*, Padua, 1656 (from Dircks 1861, as in Fig. 2.3).

'Wastebook', he made a sketch of a perpetually turning wheel powered by gravity. The wheel was half covered by a 'gravity shield': when gravity rained down on the exposed half, then that half got pushed downwards while the other half was dragged upwards, and the whole wheel began to turn. This *could* have worked if only a gravity shield could have been found…

Not one of the above machines was totally convincing, and by the middle of the eighteenth century it was becoming clear that all such attempts were doomed. In 1775, the French Royal Academy of Sciences said 'Non!', no more, and emphatically put perpetual motion in the same category as squaring the circle, trisecting the angle, doubling the cube, or searching for the philosopher's stone.

While there were always seekers after perpetual motion, there were, at the same time, always others who appreciated that this quest was vain. Most interestingly, the impossibility of perpetual motion began to be used as the basis for a variety of proofs—proofs that, with considerable hindsight, all had a bearing on the conservation of energy.

The first such proof was that due to the Flemish 'geometer' Simon Stevin (1548–1620).[10] (Stevin came from that wonderful canal-city of Bruges. He invented the land-yacht and decimal fractions.) He considered a chain draped over some inclined planes (Fig. 2.6). As perpetual motion is 'absurd', then the chain doesn't cycle round and therefore everything must be in balance (in Stevin's idealized thought experiment, the chain never catches on the pointy bits). Stevin's remarkable and simple conclusion?—the weight of chain along a given side is proportional to the length of that side. (That this had anything to do with 'forces' resolved along different directions came centuries later.)

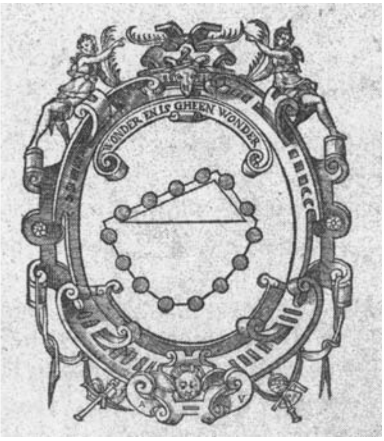

Fig. 2.6 Stevin's 'wreath of spheres' (*clootcrans*) from *De Beghinselen der Weeghconst*, 1586 (courtesy of the National Library of Australia, ANL).

A proof based on the impossibility of perpetual motion will be used again by Sadi Carnot, some 240 years later, and with enormous repercussions (see Chapter 12).

Did the sheer variety of ways in which perpetual motion could fail argue that something was being conserved? This was a question that could barely be put—let alone answered—until the middle of the nineteenth century.

3

Vis viva, the First 'Block' of Energy

Galileo Galilei

Archimedes had looked at the lever as a problem in statics. Two thousand years later, Stevin introduced motion only to ban it. Who could nowadays look at a lever and not see the motion of the parts, and what child could look at an inclined plane and not think of motion down it?

Galileo Galilei (1564–1642) was one who looked at inclined planes afresh—and brought in *motion*, mathematically, for the first time.[1]

It isn't clear whether Galileo really did carry out the legendary experiments, dropping cannon balls from the Leaning Tower of Pisa (he may have done this just for the important purpose of propagandizing and explaining by means of a demonstration). What he *did* do was realize that Stevin's inclined plane was really the same sort of problem as dropping balls from the Tower. Better than that, he realized that the inclined plane slowed everything down, so that the free-fall process was brought within experimental reach.

Let's be awe-struck and list (not in chronological order) some of the mind-shifts that Galileo brought in. First, the famous one:

(1) Bodies of different weight all fall at the same rate (have a constant acceleration).

Then:

(2) Looking at a process, at *motion*.

(3) Not just looking but measuring, in other words, mapping numbers on to physical quantities.

(4) Recognizing that an experiment is not reality, it's an idealization. (One thing that Galileo got wrong was that sliding friction can be

idealized away, but a rolling ball is different—it actually uses up energy to roll.)

The fifth mind-shift was even more momentous than the first four:

(5) 'Time' is a physical parameter.

Yes, time could be marked out by water-clocks, candle-clocks, and sundials; yes, the seven ages of man were known about; yes, the acorn takes time to grow into an oak; yes, the priest's sermon dragged on (in time) so that Galileo's attention was diverted towards the swinging of the church lamp—and so on. But that time could be put into a mathematical relationship, that it could be brought into comparison with distances travelled, that it was a 'dimension in physics'—this was new.[2]

How hard this step was is evidenced by the fact that, in his free-fall (inclined-plane) experiments, Galileo tried and tried again for *over 20 years* to find a link between one distance and another distance before he could bring himself to introduce time. When he did, he eventually found that for a body dropped from rest:

> the spaces run through in any times . . . are as the squares of those times.[3]

From this came also $v^2 \propto h$, a relationship* that would be noticed by Huygens and would become important in the story of energy.

Meanwhile, the Aristotelians (according to Galileo) were looking at free fall in a totally different light:[4]

> [the Aristotelian] must believe that if a dead cat falls out of a window, a live one cannot possibly fall too, since it is not a proper thing for a corpse to share in qualities that are suitable for the living

—a cat paradox preceding that of Schrödinger?

Incidentally, how did Galileo measure the time of fall? A water-clock would not have been suitable for such short time intervals. The Galileo scholar Stillman Drake suggests an alternative method—singing.[5] It is possible to sing to a constant rhythm, achieving a timing precision of 1/64th of a second. Drake suggests that gut lute strings (Galileo was an accomplished lute-player) were tied around the inclined plane and as the ball rolled down it would make a noise when passing over each string. With repeated trials, the position of the strings could be adjusted

* Here, v is the final speed and h is the vertical distance fallen. The symbol '$\propto$' means 'proportional to'.

until the soundings coincided with the regular beat of a song. The distances so mapped out would then be those covered in regular time intervals as the ball rolled down the plane.

This method had the advantage of high precision, but had the disadvantage of linking speed with distance rather than with time of fall. But eventually the speed versus *time* relationship in uniform acceleration was to emerge. Drake has it that this was the first time in the history of science that experiments were carried out in order to discover a quantitative relationship rather than just to demonstrate a known effect.

Mind you, Galileo was expecting the result that he finally determined; he had long realized that the Aristotelian claim that heavy bodies fall faster than light ones was wrong. First, he argued it from reason: how could a heavy composite body fall faster than its parts? Secondly, he argued it from experience: he remembered a hailstorm he had witnessed as a youth—large and small hailstones all arrived at the ground at the same time.

So Galileo realized that, on both rational and experimental grounds, the acceleration was the same for bodies of different weights—or was it? The final unravelling required the appreciation that no experiment is perfect. In his wonderful book, *Two New Sciences*,[6] Galileo answers criticisms from an Aristotelian:

> Aristotle says, 'A hundred-pound iron ball falling from the height of a hundred braccia [one braccia is 21–22 inches] hits the ground before one of just one pound has descended a single braccio.' I say that they arrive at the same time. You find, on making the experiment, that the larger anticipates the smaller by two inches…And now you want to hide, behind those two inches, the ninety-nine braccia of Aristotle.

This is likened to: 'attacking something I said [that] departs by a hair from the truth, and then trying to hide under this hair another's fault that is as big as a ship's hawser'.

Let's pause and look briefly at Galileo the man.[7] Born Galileo Galilei in Pisa, Italy, he doesn't conform to the modern caricature of a physics professor. He had a long 'marriage' to one woman, was an attentive and affectionate father to his three children, played the lute, liked his wine, and saw the commercial potential of his discoveries (the telescope and a 'geometric and military compass'). His books demonstrate a beautiful literary style, a sense of humour, and an enormous talent for polemical writing and explaining things simply. The only hint that he could be one of the most unorthodox thinkers of all time was the fact that he went against convention in never officially marrying his *de facto* wife. As regards his famous breach with the Catholic Church, it seems that the bishops and cardinals were bending over backwards to accommodate Galileo's views if only he would just shut up and not make waves. However, he kept on and on, (naively, yes) unable to accept that the church couldn't be won round by reason. He spent the last few years of his life under house arrest at his villa in Arcetri, blind but continuing to write (the *Two New Sciences* was smuggled out to Holland and published by the Elsevier family in 1638). No one knows whether he really muttered the words 'and yet it [the Earth] moves' under his breath at his recantation trial.

Galileo's introduction of 'time' was obviously crucial to the development of the concept of energy—how could something be seen to be conserved before time had been discovered? But we still haven't finished with our list of Galileo's extraordinary mind-shifts:

(6) The heavens have imperfections (the Moon has craters and the Sun has spots and rotates).

(7) The Earth moves and, in fact, orbits the Sun (supported by the phases of Venus and Jupiter's moons, further discoveries of Galileo's).

(8) The 'fixed' stars are extremely far away (follows from 7).

(9) Physics is 'written in the language of mathematics'.[8] (For example, Galileo used mathematics to attempt a solution to Xeno's paradox of how motion can ever begin: must a body accelerating from rest pass through every speed?—and how could separate increments of speed be compounded? All this with just Euclidean ratios of like quantities—no algebra, let alone calculus.)

Finally, we have the last outstanding and radical departure, possibly Galileo's single most important contribution to physics:

(10) Motion is relative.

Taking the example of the inclined plane yet again, Galileo asked what would happen if the plane was made shallower and shallower. He answers:

> a body with all external and accidental impediments removed travels along an inclined plane with greater and greater slowness according as the inclination [of the plane] is less.[9]

What happened when the plane was horizontal? The answer is given in the dialogue between a common man (Simplicio) and Galileo (Salvatio):[10]

SIMP: I cannot see any cause for acceleration or deceleration, there being no slope upward or downward.
SALV: Exactly so. But if there is no cause for the ball's retardation, there ought to be still less for its coming to rest; so how far would you have the ball continuing to move?
SIMP: As far as the extension of the surface continued without rising or falling.
SALV: Then if such a space were unbounded, the motion on it would likewise be boundless? That is, perpetual?
SIMP: It seems so to me, if the movable body were of durable material."

So here we have come to an astonishing and very non-Aristotelian result—motion on a horizontal surface, free from 'impediments' such as friction or a body that breaks up, continues forever (perpetually) and without need of any empathies, forces, or angels to push it along. Perpetual motion, *of a machine*, was looking less and less likely, but perpetual motion of an unimpeded body was a *requirement*. The English philosopher Hobbes was so bowled over that he came to visit Galileo at Arcetri in 1636 and made Galilean relativity the core of his philosophy.

There was so much that could be gleaned from this result. As the inclined plane became shallower, the acceleration was reduced until eventually it was zero—but the final speed wasn't zero. What was this speed? From $v^2 \propto h$ (see above), a body on a horizontal plane has a constant speed dependent only on the height from which it has been released. But the body itself has no 'memory' of this height, and so *it doesn't matter*. To restate this less obscurely, there is no difference between one speed and another; speed cannot be determined absolutely, only motion *between* bodies is important:[11]

If, from the cargo in [a] ship, a sack were shifted from a chest one single inch, this alone would be more of a movement for it than the two thousand mile journey [from Venice to Aleppo] made by all of them together.

Even the state of rest isn't special in any way:[12]

Shut yourself up with some friend in the cabin below decks on some large ship, and have with you there some flies, butterflies and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction... [then] have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still.

In other words, only relative motion is important and can be determined in an absolute sense. This came to be known as the Principle of Galilean Relativity and could be said to mark the start of physics. It was the very beginning of a line continuing with Descartes, Huygens, and Newton and culminating with Einstein's Principle of Relativity. Needless to say, it will be crucial in the unfolding of the concept of energy.

Galileo's principle of the relativity of motion gave very good ammunition for his arguments in favour of a moving Earth: the Earth *does* move, he said, but *we are not aware of it*.

There is, of course, one way to tell whether the ship is moving (relative to something else)—look out of a porthole and watch the view going by. For the Earth, watching the view meant looking at the stars—they were seen to rotate about the pole star every day. But was it the Earth that was spinning or were all the stars rotating in unison? Simplicity argued the case for a spinning Earth, but Galileo also considered the problem from an intriguing angle—the enormous 'power' that would be needed to have the stellar sphere—so far away, so many and such large bodies—moved instead of the Earth: 'Tiring is more to be feared for the stellar sphere than the terrestrial globe.'[13]

We have not yet mentioned the *direction* of motion. The perpetually moving body moves not only at constant speed but also in a fixed direction. Here, Galileo betrayed an Aristotelian legacy. For him, the 'horizontal' plane was only a small part of the great circle going around the centre of the Earth. He appreciated that by mathematical correctness this plane was only horizontal at one point—the point of contact of the plane with the surface of the spherical Earth. Motion away from this point was like going uphill and there would be a resistance to such motion. However, this effect could be lumped together with friction and other 'impediments to motion'. For both Aristotle and Galileo, all 'natural' motion (i.e. that which is unforced—not due to cannons, crossbows, or muscles, etc.) was either straight down towards a 'gravitating' centre or it was circular and equidistant from such a centre. A straight line extending to infinity out in space or a circle not centred on the Earth or some other 'centre' was unnatural, unthinkable.

The relativity of motion incorporated yet another mind-shift:

(11) The system is important.

For example, the ship and all its cargo constitutes a system and the shoreline is outside this system. Galileo distinguished between the cases of a body dropped from a tower and a body dropped from the mast of a ship. The position of the body relative to the bottom of the tower and the bottom of the mast was not the same. In the former case (system), both the tower and the surrounding air were carried together (on the moving Earth); in the latter case, the boat had an additional motion *through* the air.

The system can be in Venice or Aleppo or moving uniformly between the two—it doesn't make any difference. Also, the system can be scaled up or down and it doesn't make any difference. But, for any alteration, one must be careful to apply it to the *whole* system. For example, Galileo considered the case of a bone in a small animal and in a giant (Fig. 3.1), and noted that they would *not* have the same proportions (we can now appreciate that he was unable to scale up *all* aspects of the system). What constitutes the total system is important—and not always easy to identify. This will be a recurring theme in the identification of energy.

Let's now consider Galileo's work in a totally different arena—the machine. (Galileo was frequently consulted by his patron, the Grand Duke of Tuscany, about possible schemes put forward by inventors of pumps, milling machines, and olive presses.) Galileo understood that a perpetually acting machine was an impossibility: 'nature cannot be . . . defrauded by art'.[14]

However, in considering percussion the situation was not so clear, and Galileo admitted to 'dark corners which lie almost beyond the reach of human imagination'.[15] The trouble arose in comparisons between the force of a heavy dead weight and the 'energy and immense force [*energia e forza immensa*]'[16] of a much lighter but moving weight, such as the blows of a hammer. The hammer could act as a pile-driver, whereas in gently removing and replacing the heavy dead weight, nothing happened. This problem had to wait for the advent of Newton and his mechanics for its resolution.

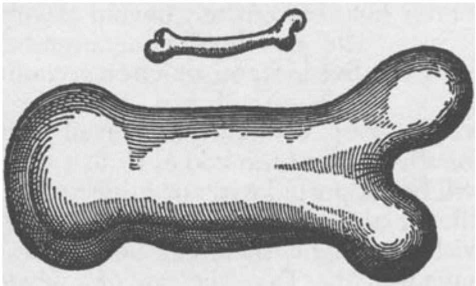

Fig. 3.1 Bone in a small and a large animal, from Galileo's *Two New Sciences*, 1638 (with the permission of Dover Publications, Inc.).

René Descartes

In histories of physics, René Descartes (1596–1650) is hardly ever given his due. Somehow his famous 'Cogito ergo sum' ('I think, therefore I am') has resulted in his being hijacked by the philosophers. However, Descartes was just as motivated to develop his theory of the physical world (*res extensa*) as his theory of the mental world (*res cogitans*). These two domains—extension and thought—were the only ones that Descartes' 'method of doubt' allowed him to be certain about.

Descartes had a life-changing experience that made it clear to him what his mission in life was: to be a philosopher and to construct a complete system of knowledge. Apparently, while shut up in an oven (a sort of boiler-room), he had a day of solitary thought and a restless night of dreams. It became clear to him what the correct path was. He had to start from no prior knowledge, no beliefs, and then, as it were, re-skin the onion, admitting only those layers of knowledge, one by one, about which there could not be any doubt. Another interesting anecdote about Descartes tells of his antipathy to getting up early. He never started work before 11 in the morning. When he went to Sweden to tutor Queen Christina in 1650, she insisted that lessons start at 5 a.m. This, combined with the harsh Swedish climate, was too much for Descartes and he caught a chill and died.[17]

If the physical domain was characterized only by the property of extension, where did matter come in? Descartes answered this by considering that the chief property of matter *was* extension. He also argued this the other way around: if there was no matter, then there was also no extension. In other words, a void or empty space was an impossibility. This seems very strange to us, even more so when we recall that Descartes invented the way of representing space still used today—the Cartesian coordinate system. Perhaps because he *had* invented it, he couldn't believe that it corresponded to something real.

The straight axes of Descartes' coordinate system extended to infinity. For the first time, the universe was of infinite extent. The realm of the stars in Copernicus' celestial system was much further away than it had been in Ptolemy's, but all the celestial bodies still moved in *finite* perfect circles. This was also true in Galileo's universe. But Descartes had a more abstract concept of motion altogether. Despite his infinite universe, he thought that only local motion made any sense. Also, he didn't accept Galileo's division of motion into natural and unnatural (forced) kinds and replaced them by a single, more general and abstract kind: rectilinear (in a straight line), unaccelerated, unbounded, and unconnected with any centres of activity. This intellectual leap was all the more impressive in view of the fact that such rectilinear motion was never to be observed in practice, as Descartes' universe was completely filled with matter (a plenum) and so unobstructed, rectilinear motion could never actually take place.

The infinity also operated in the reverse direction: matter was infinitely divisible—the smaller parts could always be subdivided into even smaller parts. Space was completely filled with this continuum of matter and so atoms didn't exist. This was consistent with Descartes' banishing of the void: if atoms existed, then there would have to be empty space between those atoms. Descartes did, however, posit three types of matter.[18] The first element (fire) was 'the most subtle and penetrating liquid'; the second (the element of air) was also a subtle fluid; and the third (the element of earth) was made from bigger, slower particles.

Common to all these forms of matter was the fact that matter was inert, had extension, and could be in motion. However, as space was already completely full, this motion could only occur through some matter displacing other matter in a continuous loop. This loop or vortex motion was made to account for the orbits of the planets and for gravity at the Earth's surface. The trouble was that it didn't reproduce Galileo's free-fall results.

However, the important point to notice is the leap made by Descartes in going from matter moving in closed whirlpools, bordering and constrained by other whirlpools, to the motion that matter *would* have if it wasn't so constrained. Descartes considered this hypothetical, 'free' motion (we now say 'inertial') to be of constant speed and rectilinear. His first two laws of nature were as follows:[19]

> That each thing remains in the state in which it is so long as nothing changes it.
>
> That every body which moves tends to continue its motion in a straight line.

He appreciated that the important question was not what would keep a body in motion, but what could cause that motion to cease:

> we are freed of the difficulty in which the Schoolmen [the Scholastics] find themselves when they seek to explain why a stone continues to move for a while after leaving the hand of the thrower. For we should rather be asked why it does not continue to move forever.[20]

Descartes appreciated that motion was relative and that a state of rest had no special status compared to a state of motion. He may well have exploited this relativity in order to conceal his adoption of the Copernican system with its requirement for a moving Earth (Descartes had taken careful note of Galileo's troubles with the Church):

> The Earth is at rest relative to the invisible fluid matter [Descartes' second element] that surrounds it. Even though this fluid *does* move [around the Sun] and carries the Earth with it, the Earth does not move any more than do those who as they sleep are carried in a boat from Calais to Dover.[21]

The original source of all matter and motion was God. As God and His creations were immutable, the total quantity of matter and of motion could not change, although particles could acquire or lose motion after colliding with other particles. Indeed, this was the fundamental mechanism of change in Descartes' universe. All physical phenomena (e.g. gravity and magnetism) and all qualities (colour, smell, etc.) were to be explained solely in terms of collisions between bodies. No occult forces, attractions, empathies and antipathies, action-at-a-distance, and so forth were to be admitted. This was the seventeenth-century's 'mechanical philosophy': it had many followers (such as Boyle and Gassendi), but Descartes was its chief exponent.

The conserved quantity of motion was defined by Descartes as the 'mass' of a body times its 'velocity'. 'Mass' was an umbrella term including the size and shape of the body. 'Velocity' was what we would now call speed, as the direction ('determination') of a body did not contribute to its quantity of motion.

In the mechanical philosophy, collisions between moving bodies were to account for every process in the physical world. A theory of collisions was therefore of paramount importance to Descartes and this was one of the few occasions when he attempted a quantitative treatment. His guiding principle was that the total quantity of motion of two bodies after collision should equal the total quantity of motion before collision. This has a familiar ring to us—it sounds like the law of conservation of momentum—but Descartes' conception of motion was such that the quantity of motion was always positive: there was no change in sign following a reversal of direction.

(A change of direction did not presuppose a 'force' according to Descartes. This allowed changes of direction to be used in a peculiar way. The two domains of Descartes' world—thought and extension—were distinct, but how did these domains interact? For example, how does a person will his hand to open and close? Descartes postulated that minds could affect just one aspect of moving bodies—their direction of travel. These redirections occurred in the pineal gland sited in the brain. Only human brains could do this; all other animals had no minds and were automata. There was much correspondence on this topic between Descartes, Gassendi, and Hobbes.)

In the cases of perfectly elastic or inelastic collisions between identical bodies, Descartes' rules gave results identical to the accepted ones. However, in most cases his rules did not agree with experiment. The most glaring mismatch occurred in the case of a collision between a small body moving at constant speed towards a larger body at rest (rule 4). Descartes maintained that the larger body could never be moved by the smaller one, no matter how rapidly this smaller one approached. Of course, Descartes *did* realize that this didn't work out in practice and so he developed the perfect fudge: his rules were only for the collisions between two bodies in isolation—that is, in a vacuum—but the vacuum didn't exist. In reality (argued Descartes), there were innumerable collisions between gross bodies and the matter making up the invisible fluid medium (Descartes' second category of matter) and these collisions would be just such as to account for the differences between theory and experiment... (Descartes tried to use a similar argument to explain why his vortex theory of gravity didn't accord with Galileo's experiments on the uniform acceleration of freely falling bodies.)

As well as his collision theory, Descartes also gave a new analysis of two other specific problems.

First, he examined circular motion and realized that this was made up of a particle's rectilinear motion along a tangent to the curve and another component at right angles to this. This other component was the particle's tendency to recede from the centre of rotation—what Huygens would later refer to as a centrifugal tendency.

Secondly—and of great relevance to energy—Descartes investigated the fundamental laws behind the so-called simple machines (the lever, pulley, inclined plane, and so on). He started with a general principle that certainly has a foretaste of an energy conservation principle:[22]

> the effect must always be proportional to the action that is necessary to produce it...

Descartes saw that despite their various methods of working the 'action' of all the simple machines could be generalized to just one case:

> the action... is called the force by which a weight can be raised, whether the action comes from a man or a spring or another weight etc.

and he was able to propose a quantitative measure for such action:

> The contrivance of all these machines is based on one principle alone, which is that the same force which can raise a weight, for example, of a 100 pounds to a height of two feet can also raise one of 200 pounds to the height of one foot or one of 400 to the height of a half foot, and so with others, provided it is applied to the weight.

Descartes' measure for action is identical to our measure for work (against gravity) and as such is the very start of our quest for the concept of energy.

Descartes' legacy was almost as boundless as his universe: the merging of geometry and algebra, the mechanical philosophy in which animistic and occult influences were given short shrift, the generalizing of the concepts of 'free' motion and of circular motion, and, of most particular importance to our theme, the emergence of two measures—the 'quantity of motion' (roughly mass × speed) and the 'action' of simple machines (weight × height).

However, Descartes' world-view was empirical only to the extent that the facts of nature were to be taken as a given—there was no need to actually carry out experiments (although he did do animal dissections and some experiments in optics). Physical knowledge could be built up by reason alone ('my physics is nothing but geometry'[23] he once wrote) and ultimately his physics was flawed in this respect.

Christiaan Huygens

Christiaan Huygens (1629–95) had the misfortune to fall between Galileo and Descartes on the one hand and Newton and Leibniz on the other—his greatness would otherwise have been more readily apparent. Today, Huygens is famous to physicists for his wave theory of light, to astronomers for discovering the rings of Saturn, and to clockmakers (maybe) for his magnum opus, the *Horologium oscillatorium*.[24]

Huygens was born in The Hague into a cultured family of Dutch diplomats serving the house of Orange. Descartes was a frequent visitor and Huygens was brought up immersed in the prevailing Cartesian orthodoxy. His tutor was therefore shocked to find that Huygens was working on some errors in Descartes' theory of collisions.

The results of Huygens' studies on collisions were collected into a treatise, *De motu corporum ex percussione*,[25] in 1656, and the most important theorems were presented to the recently formed Royal Society in London in 1668, in response to a call for papers specifically on this topic. John Wallis (Professor of mathematics at the University of Oxford) and Sir Christopher Wren (the mathematician and architect) were the other respondents.

Apparently, Huygens and Wallis shared a long stagecoach journey[26] during Huygens' visit to England from Paris, where he was working at the even more recently formed Académie Royale des Sciences: one can only speculate as to the conversation and wonder which language it was conducted in. Despite frequent wars between France, Holland, and England during the latter half of the seventeenth century, a free interchange of ideas continued between the natural philosophers of the day. However, Huygens was guarded about his work and often coded his results as anagrams or numerical ciphers. He lived at Louis XIV's expense in an apartment at the Bibliothèque Royale. In later years, while Huygens was on a visit to The Hague, his patron, the Minister Colbert, died and the tide turned against protestants. Huygens didn't return to Paris, but stayed on in Holland for the rest of his days.

Getting back to the *De motu*, Huygens put forward a series of both hypotheses and propositions, and set out to refute Descartes much in the manner of a proof by Euclid or Archimedes:

Hypothesis I, A body will continue in a straight line with constant speed if its motion is not opposed.

Hypothesis II, Equal bodies approaching with equal speeds will rebound with their speed of approach.

Hypothesis III, For any collection of bodies, only their motion relative to one another makes any difference.[27]

The first hypothesis reasserted the principle of free or unforced (what will, post Newton, be called inertial) motion inherited from Galileo and Descartes. His second hypothesis appealed to inherent symmetries and the third hypothesis reasserted Galileo's profound insight on the relativity of motion. The next hypothesis was not of universal scope like the others, but Huygens found that he couldn't determine all possible outcomes unless this particular case was specified:

Hypothesis IV, A big body hitting a smaller one at rest will lose some of its motion to the smaller body.

Finally, in the fifth hypothesis, Huygens managed to generalize the scenario of the second hypothesis (equal bodies, equal speeds) to the case of unequal bodies:

Hypothesis V, Two bodies whatever [*quelconques*] collide: if one conserves all its motion then so does the other one.

This hypothesis stressed the idea of conservation rather than symmetry. Huygens was so convinced of its importance that in the original manuscript he underlined it with a wavy line.

Having established this base of hypotheses, Huygens had sufficient theoretical weapons in his armoury to begin his attack on Descartes. The rules of impact established by Descartes disagreed with experiment. The most conspicuous error was rule 4, that a fast-moving small body could never shift a larger body at rest. Huygens was of a different ilk to Descartes—no fudge could ever conceal an outright clash with observations. But Huygens' objections went deeper. Descartes' 'quantity of motion' (*mv*) had to maintain an absolute constant value, while Descartes still adhered to the relativity of motion: but how could motion be both relative and absolute at the same time? There was no doubt in Huygens' mind that of the two principles it was the relativity of motion that had to be retained.

Although Huygens was a Cartesian, his outlook and inspiration derived from Galileo's works. Galileo had asserted what is now referred to as Galilean Relativity, that there is no absolute state of motion, and Huygens took this on board—quite literally: he envisaged a collision on board a smoothly coasting canal boat and then viewed both from this

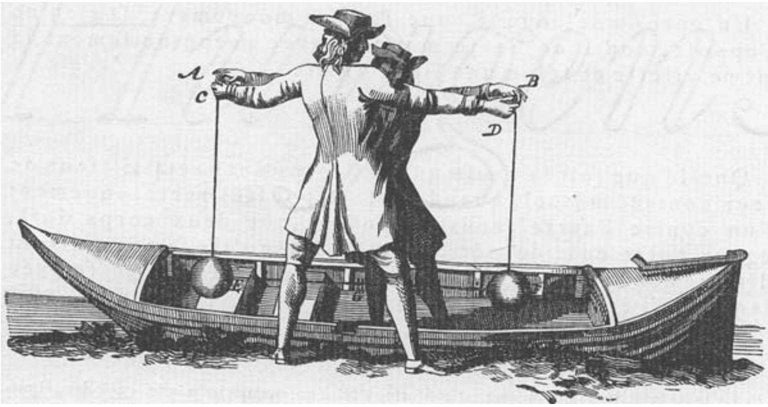

Fig. 3.2 Huygens' thought experiment in *De motu corporum ex percussione*, 1656 (published by Martinus Nijhoff NV).

boat and from the banks of the canal. The frontispiece to the *De motu* (Fig. 3.2) shows the imagined experimental set-up.

The experiment was essentially a collision between two pendulum bobs ('bodies'). The bodies were taken to be spherical and infinitely hard, making the collision perfectly elastic—in theory at least. The experiment is not as feasible as it looks. Apart from external impediments (air resistance, torsion in the cords, and so on), how would the bodies be given specific starting speeds and how would the speeds after impact be measured? This was almost certainly a thought experiment rather than a real one. (Another contemporary metaphor may have been 'billiard-ball collisions' as the game of billiards was popular at this time and familiar to Huygens.[28]) Nevertheless, Huygens' thought experiment was used in a new way. This was the first time that Galilean Relativity was put to a *quantitative* test—two different viewpoints (reference frames) were compared.

On board the canal boat a small body collides with a larger body at rest. If the small body has the same incoming speed, v, as the canal boat but in the opposite direction, then for an observer on the bank it will appear as a collision between the larger body moving with speed v and the smaller body at rest. Descartes had different outcomes for these two viewpoints, but Huygens insisted that, by Galileo's principle of relativity, the outcomes should be the *same* (there was, after all, only one experiment).

Huygens had managed to correct the error in Descartes, but he wanted to go further and discover the overriding principles governing all collisions. He continued with propositions rather than hypotheses. First, he proposed that the relative speed of approach equals the relative speed of recession (Proposition IV) and then that an elastic collision will always be reversible (Proposition V). By Proposition VI he was ready to propose that Descartes' 'quantity of motion' (*mv*) is not always conserved. Interestingly, although Huygens managed to show that direction was the ingredient missing from *mv*, he didn't elevate this to an axiom, a hypothesis, or even a proposition. He still thought of motion in Cartesian terms; it was to him always a *positive* quantity.

Huygens was still not satisfied that he had reached a general principle from which all the rules would follow—until he arrived at Proposition VIII:

> for two bodies approaching each other with speeds inversely proportional to the magnitude of the body, each will rebound with a recession speed equal to its approach speed[29]

—by 'magnitude', Huygens meant roughly 'mass'.

Huygens realized that he had at last found what he was looking for—a proposition that preserved the symmetry of Hypothesis II, but that was applicable to the more general case of unequal bodies and unequal speeds.

All these hypotheses and propositions appear rather bloodless to the modern reader—no body ever moves because it has been *hit*. Huygens tried to avoid all mention of force and instead stressed symmetries and conservation. It should be borne in mind that all this *precedes* Newton and the latter's definitions of force, mass, and momentum and the 'game plan' provided by today's Newtonian mechanics.

Why did Proposition VIII work? Huygens recognized that in the special scenario where the bodies have speeds that are inversely proportional to their 'magnitude', then the centre of gravity of the total system of bodies remains fixed. While the proposition singles out a special viewpoint (that in which the centre of gravity is at rest), the principle of relativity guarantees that any results found from *this* viewpoint can be applied to all other viewpoints.

The centre of gravity was a concept that had been known about since antiquity and, as we saw in Chapter 2, was taken by Archimedes to be self-evident. But why was Huygens so sure of the truth and importance of Proposition VIII? He recalled the work of Evangelista Torricelli

(1608–47; famous as the inventor of the barometer) and his result that when two bodies are linked together and constrained to move only in the vertical plane, no movement freely undertaken by these bodies can raise their common centre of gravity. Huygens generalized Torricelli's result to any number of bodies moving any which way and realized that 'the common centre of gravity of bodies cannot be raised by that motion of the bodies which is generated by the gravity of those bodies themselves'.[30] In other words, levitation is impossible (in the same way that you can't raise yourself off the ground by pulling up on your hair). So, the centre of gravity must remain at the same level—but more than that, its motion must be 'inertial' (must satisfy Hypothesis I). Huygens finally arrived at the following[31]

ADMIRABLE LAW OF NATURE... It is that the common centre of gravity of two or three or however many bodies you wish continues to move uniformly in the same direction in a right line before and after their impact.

In other words, whatever the details, the mass acts as if it is one total mass concentrated at the centre of gravity. This centre of gravity continues to move perpetually, sublimely unaware of what the constituents are up to.

But what is the relevance of prohibiting the centre of gravity from undergoing any change in *height* when Huygens was considering bodies moving only in a *horizontal* plane? Huygens took a leaf out of Galileo's book (the *Discourses on Two New Sciences*), where bodies are dropped from certain heights and then fall freely until deflected into horizontal motion. Galileo had found that v^2 is proportional to h. All that was needed was a way of converting the vertical motion into horizontal motion and vice versa. Huygens' frontispiece is suggestive—the bodies are attached to cords and can therefore be released from prescribed heights. In a footnote Huygens has a tiny diagram that shows another possibility—elastic bands (or whatever the seventeenth-century equivalent would have been) at a 45 degree angle to the vertical and acting as deflectors.

In reverse too, a body already on the horizontal and moving uniformly with speed, v, when deflected upwards will have its speed retarded until finally, at the *same* height, h, the motion is exhausted (used up). Galileo had asserted this reversibility (assuming no impediments) and Huygens saw that this *must* be so, as otherwise a perpetual motion would result.

Thus, ultimately, it was the inadmissibility of perpetual motion that brought in the requirement that the centre of gravity maintains... a perpetual motion.

This contradiction is resolved when it is understood that the banned perpetual motion refers to a 'perpetual *process*' and not to a 'perpetual *inertial* motion'. It was this impossibility of a 'perpetual process' that Huygens saw as the real physical explanation behind all the rules of collision.

(Before we jump to conclusions ahead of the times—the seventeenth century—it is worth noting that Huygens still hadn't ruled out the possibility of a perpetual motion by *non*-mechanical means. For example, he wrote of a clock invented by Robert Hooke (1635–1703):

> [a] clock invented by Monsieur Hook [*sic*] . . . moves with the aid of a loadstone: if it uses no other motor, this would be a type of perpetual motion and it would be an admirable invention.[32])

So Huygens, who wanted to keep all considerations of forces out of the picture, had to allow gravity as the force that, in collisions, gave the bodies their initial speeds and soaked up their recoil speeds. (Huygens was never able to accept Newton's Law of Universal Gravitation because of the occult requirement for action-at-a-distance.) In this way, height came into a scenario that was totally on the horizontal. Changes in the height of the centre of gravity were not allowed in order to rule out perpetual motion (of a process), while perpetual motion of the inertial kind was endorsed.

This seems like reasoning a long way round, but it had a very interesting by-product. For the height of the centre of gravity to remain constant, the mathematics showed that another quantity had to stay constant: the total mv^2 before and after the collision. This was the first time that mv^2 made an appearance in physics. Huygens merely noted that it was a conserved quantity and acknowledged its consequent utility as a predictor of the final speeds after collision.

In Huygens' later work on circular motion (*De vis centrifuga*, 1659), Huygens finally succumbed to the idea of force and coined the term 'centrifugal' force to describe the tendency of a circling body to flee the centre of rotation. He found that this force increased with the speed of rotation and decreased for a larger radius, but what about the mass? 'Mass' was still a perversely insubstantial thing and Huygens always referred evasively to 'equal', 'unequal', or even 'whatsoever' bodies ('*corps quelconques*'). ('Mass', as we know, will, some 250 years into the future, be intimately linked with energy.)

However, in connection with centrifugal force, Huygens contrived another thought experiment: a homunculus sitting on the rim of a giant wheel and holding a mass or bob on a cord. As the wheel rotates, the bob pulls away from the centre of rotation. Substituting greater and greater masses for the bob, Huygens noted that the centrifugal force increases proportionately (the bob pulls away more and more strongly). The parallel with gravity was not lost on Huygens. Professor Henry Crew[33] notes that 'This is the first time that a distinction between mass and weight appears in physics'—yet another first for Huygens. (That mass can exhibit weight merely because of its acceleration will have enormous repercussions in the hands of Einstein.)

To summarize, Huygens was the first to apply Galilean relativity in a quantitative way. He was also the first to bring out a distinction between 'mass' and 'weight'. Most crucial of all for energy, he showed that a certain quantity—a mere algorithm ('*calculatrix*') in the case of collision experiments—was conserved. This was the quantity mv^2.

Isaac Newton

In 1687 Isaac Newton (1643–1727) published his great work, *Philosophiae naturalis principia mathematica*[34] (*The Mathematical Principles of Natural Philosophy*), always referred to as 'the *Principia*'. As claimed in the title, the book set out the founding principles of physics: *it is the most influential book ever written*. I am not going to celebrate Newton's achievements but, rather, comb his work for anything of relevance to 'energy'.

Newton starts off with defining mass for the first time. (Huygens sidesteps the issue by referring to a body '*quelconque*'.) He defines 'mass' as being equal to the volume of a body times its density, where 'density' and 'volume' are taken to be understood. Significantly, the definition is in the measure—a recurring approach in Newton. He performs careful experiments (timing pendula with bobs made from hollowed-out spheres of gold, copper, silver, and wood) and finds that the weight of a body is proportional to its mass—also very significant, but that's another story, for the time being. He asserts that matter ('mass') is made ultimately from infinitely hard atoms that are indivisible, impenetrable, and capable of motion. The quantity of matter then boils down to a head-count of the atoms.

Masses or bodies move about in empty space—a complete departure from Descartes' plenum. (Newton bought Mr Carte's *La Geometrie* at a fair in Sturbridge, Cambridge, and tossed it aside until persuaded to have another look by his tutor, Isaac Barrow.[35]) But Newton does follow

Cartesian ideals in a number of ways: the physical world is materialistic (as opposed to, say, animistic or psychic) and matter is passive (Newton calls it inert) and unaffected by its own motion. This passivity or inertia is exemplified by the law of 'inertial' motion already discovered by Galileo and enunciated by Descartes and Huygens. Newton gives it prominence as his First Law of Motion:

LAW I: Every body perseveres in its state of rest, or of uniform motion [meaning at constant speed] in a right line, unless it is compelled to change that state by forces impressed thereon.[36]

Newton defines such a uniformly moving, constantly directed body as having a '*vis insita*' or '*vis inertia*'—an internal 'force' that keeps the body moving. But, in contrast to Descartes, for Newton this is emphatically not the *true* force. First, it can be cancelled out by motion in the opposite direction (Newton sticks to his mathematical guns on this) but, more importantly, it is only ever a relative quantity, dependent on the observer's point of view.

Uniform motion is relative but one thing can be determined absolutely and that is a difference between earlier and later motions, a *change* in motion.

But what causes the change in motion? Newton looked at Huygens' bumper-car kinematics (excuse the anachronism), obeying all the symmetries of collisions, and realized that this would still never be *enough* to account for everything, for all the phenomena in the physical world. Huygens had ruled out 'perpetual motion', but what had he ruled *in*? Some new entity—force or '*vis*'—was required:

LAW II: The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed.[37]

Thus, Newton's First and Second Laws, taken together, showed that if there was *no* change in motion, then there was *no* force; if there *was* a change in motion, then there *was* a force; and if there *was* a force, then there *would be* a change in motion.

'Alteration of motion' was ambiguous. It meant 'rate of alteration of motion'—or *acceleration*. Such a denomination was only possible because of a revolution in mathematics, Newton's calculus (although Newton didn't actually use his calculus in the *Principia*). Despite the fact that Newton had written 'alteration of motion' instead of 'rate of alteration of motion', in all his examples and calculations a certain time interval was assumed, either implicitly or explicitly, and so no errors arose.

The new 'motive force' had magnitude (proportional to the acceleration) and direction. For the first time the force did not reside *in* the body but was external to it, impressed *on* the body. Also, and for the first time, the effect of this true force was not to maintain motion but to *change* it: even a decrease in the speed was the result of a force. Also, and again for the first time, a change in direction required a force—so the perfect circular motion of Galileo was no longer unforced or inertial. The effect of the force was not merely qualitative but quantitative: the change in the body's motion was proportional to and in the same line as the impressed force. Finally, two or more forces could be added by the 'parallelogram law' for the addition of directed quantities (this is immediately given as Corollary I, after the Laws of Motion).

The amazing thing about Newton's Second Law is not so much the invocation of an invisible, almost magical, 'force', but the fact that it has just one effect—to cause an acceleration (it could, for example, have left the motion unaltered but caused the mass to swell, or it could have caused a second-order change in the velocity and so on). Also, this one effect is then to account for *everything* in the physical world (waves crashing on the beach, the trajectory of a pebble thrown into the sea, a fly walking on water, the path of the Moon—and so on).

Strange to say, Newton cited Galileo as the source of the Second Law, presumably because of the latter's discovery of the constant acceleration of bodies in free fall. But acceleration is not always so evident as the consequence of a force. Take the modern-day exemplar of Newtonian Mechanics, the 'simple' case of billiard-ball collisions: where is the force and where are the accelerations? The bodies are not seen to slow down before impact or to speed up afterwards. In fact, the only evidence of anything happening is a change of direction. But Newton, armed with intuition and with mathematics, understood that this too counts as acceleration.

This is the true import of the Laws: motion is relative, but *happenings* are absolute, and need a cause. The First Law defines the simplest case—nothing is happening; uniform motion doesn't count as a happening (the only thing that different observers can be guaranteed to agree on in this case is the number and mass of the bodies). Anything more complicated, such as a collision between two bodies, constitutes 'something happening'. All observers will agree that this collision and rebound takes place even though they may disagree on velocities and directions.

The absolute acceleration opened the way for the possibility of an absolute force, as Newton required. However, this absolute acceleration seemed to rely on the identification of an absolute space that was itself absolutely unaccelerated. This problem was not lost on Gottfried Leibniz (see below), who was all too ready to pick through Newton's masterwork and find fault with it. The nineteenth-century physicist and philosopher Ernst Mach (1838–1916) also criticized Newton's absolute space and time. There was no absolution! Newton had managed to wrest something absolute from out of the shifting sands of relativity and then was criticized for his need for an absolute space and time.

However, this turned out to be mostly a small red herring—let's say, a red anchovy—as Newtonian Mechanics, in distinction to Newton, does not, for the most part, require space and time to be absolute. In other words, only *intervals* in space and *intervals* in time occur in the formulae—there are no absolute designations such as '14 October 1066' or 'the battlefield at Hastings'. It is for this reason that Newtonian Mechanics is so incredibly successful, explaining everything from billiard balls to the trajectories of spaceships. It does begin to fail at very high speeds, but not because of those speeds *per se*. It fails just because a disallowed absolute—the speed of light—is being approached. (Later, in Einstein's Theory of Special Relativity, we shall find that the separate space and time intervals are not enough to save Newtonian Mechanics.)

Another philosophical problem with the Second Law is that it appears to be circular—defining mass in terms of force and force in terms of mass. (Actually, 'mass' isn't even mentioned. We need to substitute 'rate of change of motion *of a body*' for 'change of motion'.) In fact, the Second Law would be circular—and devoid of empirical content—if applied to only one case. However, it is a statement about the physical world and it predicts the results of countless experiments, such as applying the same force to bodies of different mass, or to bodies of the same mass but made from different materials, and so on (we then only need to agree on the notion of 'sameness').

Mach takes Newton's Second Law as amounting to a definition of mass as 'reciprocal acceleratability'. However, one can continually pass the buck round when it comes to the attribute 'most-fundamental-and-irreducible-entity'. It is true that 'mass' could take on some of the burden of causing accelerations instead of leaving it all up to 'force' (and we shall see that this *is* what happens in d'Alembert's Principle (Chapter 7) and Einstein's Theory of General Relativity (Chapter 18). There is no doubt, though, that Newton's intention was to consider mass as inert and to have all the changes as being due to the newly introduced agency of activity—force.

Except in the one case of gravity, Newton did not specify the exact mathematical form of this new entity, force. He did identify many examples of force such as stretched springs, magnetic attractions, cohesion of bodies, centripetal forces, and electrostatic forces, as well as gravitational attraction. The *same* force occurs when the conditions are the same: for example, a given spring stretched to a given extent, attraction to a particular piece of loadstone, or a gravitational attraction between two specific bodies.

However, Newton did introduce some constraints. Forces arise from some material source—there are no disembodied forces or attractions to a point in space (actually, this did occur on a few occasions and Newton wasn't very happy about it). Forces have both direction and magnitude (in modern language, they are vectors) and multiple forces are added in the same way as vectors are added (Newton sticks to his mathematical guns again). Newton's forces may be attractive or repulsive and they always act along the line joining the bodies. Finally, if the external force arises from just one body, say A, and acts on one other body, B, then Newton's Third Law of Motion imposes the constraint that B must exert an equal force back on A:

> To every action there is always opposed an equal reaction; or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.[38]

Newton goes on to give examples of this law:

> If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the stone.[39]

The Third Law leads to the same results (Newton calls them corollaries) as Huygens' rules of collision: the total quantity of motion (in modern terms, the momentum) is conserved, and the common centre of gravity is unaffected by the motions of the individual bodies. But if the finger presses the stone more strongly, the stone will press back more strongly while nothing moves and so the total momentum remains at zero. Newton is therefore asserting something more than just momentum conservation; he is asserting an equilibrium between the action and reaction forces. In the uniform and unchanging motion of the common centre of gravity both Newton and Huygens saw the evidence of a larger truth—the absurdity and impossibility of perpetual motion. In the Scholium following the Laws of Motion, Newton gives an example to demonstrate this:

> I made the experiment on the loadstone and iron. If these, placed apart in proper vessels, are made to float by one another in standing water, neither of them will propel the other; but, by being equally attracted, they will sustain each other's pressure, and rest at last in equilibrium.[40]

If the attraction of the loadstone on the iron had been greater than the attraction of the iron on the loadstone, then the iron would accelerate towards the loadstone '*in infinitum*' and an absurd perpetual motion would result.

Newton went on to apply this same *reductio ad absurdum* argument to a thought experiment in which the Earth is partitioned into two unequal parts. James Clerk Maxwell (1831–79), the great nineteenth-century physicist, imagined the 'parts' as a mountain and the rest of the Earth[41] and found that if the Earth attracted the mountain more than the mountain attracted the Earth, then the ensuing perpetual acceleration would be only a very small perturbation of the usual Earth–Sun orbital motions. Therefore, claimed Maxwell, the Third Law was not to be justified on experimental grounds but stood, rather, as a reaffirmation of the First Law.

However, it seems to me that Newton put forward his Third Law on both *a priori* grounds *and* experimental grounds—it was a law that had to apply (he thought) and a law that was borne out in practice. We shall find that there *are* instances where the Third Law fails (see Chapter 18).

(Mach points out that the Third Law really amounts to a definition of mass. Mass comes out as 'reciprocal acceleratability'—again. As this has already been shown to follow from the Second Law, Mach makes the point that the Third Law is adding nothing new. But…economy of description isn't everything. Mach's view doesn't bring out the symmetry in the forces, Newton's main intention. Also, acceleration is not so 'easy' and primitive a concept as Mach supposes. It has become so only through familiarity—with cars and smooth roads, computer simulations, and so on.)

Newton goes on (in the Scholium)[42] to give examples of his Third Law as applied to the simple machines (the balance, pulley, wheel, screw, and wedge) and sums up their purpose as being 'to move a given weight with a given power'. This was seized on by Thomson and Tait (see Chapter 16) in their textbook on physics written almost 200 years later as proof that Newton had after all discovered the concepts of work and mechanical energy. (Tait shows up again in a reprehensible case of chauvinistic prejudice against Mayer; see Chapter 14.) However, it is clear that while Newton did consider perpetual motion to be an absurdity, he attached no significance to quantities such as 'mv^2' or '$\int F\, ds$' (later linked to the concepts of kinetic energy and work, respectively), even though these quantities did crop up in his calculations and later works.

When Newton uses the Third Law to analyse examples of simple machines[43] he strangely applies the action and reaction forces to the *same* body rather than to different bodies. As Professor Home says,[44] this would today be considered a schoolboy howler, but it seems unlikely that Newton is applying his own law incorrectly! Home suggests that the answer lies in a less restricted use of the terms 'action' and 'reaction' by Newton than that used today. For example, in the *Opticks*,[45] Queries 5, 11, 30, and 31, [my italics], Newton asks whether:

the Sun and fix'd Stars are great Earths vehemently hot, whose heat is conserved by the greatness of the Bodies, and the mutual *Action and Reaction* between them, and the Light which they emit; [Query 11]

Bodies and Light *act* mutually upon one another; [Query 5]

and

active principles, . . . are the cause of gravity, . . . and the cause of fermentation. [Query 31][46]

In Query 30, Newton speculates further:

Are not gross bodies and light *convertible into one another*?

In other words, Newton is giving 'action' and 'reaction' a broader meaning beyond just the accelerative force implied in his Second Law of Motion. This broader conception of 'active principles' coupled with the suggested interconvertibility between light and bodies looks forward in some respects to our modern idea of energy.

However, Newton's analysis of inelastic collisions was one occasion where his force-view led to a radically different result from the energy-view to come. In Query 31 of the *Opticks*, Newton considers a collision in a vacuum between two identical absolutely hard or perfectly inelastic bodies and finds that:

They will by the laws of motion stop where they meet, and lose all their motion, and remain in rest.[47]

The total momentum is conserved and the action and reaction forces are equal and opposite, but there is an absolute loss of 'motion' (we would say 'energy'), and Newton sanctions this loss. Thomson and Tait make no comment on this. Newton goes on to say that if such a loss of motion is *not* to be allowed, then the impenetrability of the bodies will have to be sacrificed instead. This is discussed further in Chapter 7.

What of machines and engines? Newton refers to them briefly in the Scholium (as mentioned above), but he is strangely uninterested and in the entire corpus of his work, the *Principia* and elsewhere, there is no other treatment of them.[48] This is in contrast to many of his contemporaries (for example, Huygens proposed a gunpowder combustion engine).

Newtonian Mechanics occurs in an idealized frictionless world, mostly of celestial objects. When friction *is* treated quantitatively it is treated as the resistance felt by a body moving through a liquid rather than as a loss in the efficiency of a machine.[49] All the various microscopic (frictional) effects are lumped together as one net force ('the resistance') and its effects on one given body are calculated.

In the celestial domain friction—and also concepts such as 'work' and 'kinetic energy'—aren't very relevant. Who, today, is interested in calculating the kinetic energy of Mars? However, Newton did appreciate that even for these heavenly bodies motion would be lost to air resistance. He estimated that at a height of 228 miles above the surface of the Earth, the air would be '1,000,000,000,000,000,000 times rarer',[50] but that motion *would* be lost eventually ('motion is much more apt to be lost than got, and is always upon the decay'[51]) and the planets would cease orbiting if there were no active principle to bring about a 'reformation'[52] of the heavenly order. Once again, Leibniz leaped into the fray, claiming to be outraged that Newton's clockwork universe needed God to wind it up occasionally.

We have seen that the force-based view and the energy view to come have many parallels. Newton's conception of force was more than something just to explain dodgem-car kinematics, but to explain everything—attractions due to gravity across vast distances and much stronger attractions (he even estimated how much stronger) occurring at very short ranges and accounting for chemical effects, the cohesion of bodies, the crystal structure of solids, and so on. Thus force was given a cosmic role—a similar sort of cosmic role that energy would acquire. Considering also Newton's suggestion of the interconvertibility of light and matter takes Newton even closer to the modern idea of energy. In the case of perfectly inelastic collisions, however, Newton was prepared to accept an absolute loss of 'motion'. Newton was the greatest physicist who has ever lived—or maybe that was Einstein, such comparisons are 'incommensurable'—so does it really matter if Newton let this one, energy, go by?

Gottfried Leibniz

The term 'polymath' always comes up in connection with Gottfried Wilhelm Leibniz (1646–1716)—the Germans refer to him as a *Universalgenie*. Born in Leipzig, he advanced philosophy, logic (he pretty well invented mathematical logic, but his work was lost until uncovered by Bertrand Russell in the twentieth century), law, theology, linguistics, mathematics (he discovered the infinitesimal calculus later than, but independently of, Newton), physics, and other areas. He was satirized mercilessly as Dr Pangloss in Voltaire's *Candide*, chiefly for maintaining that the world we live in is the best of all possible worlds. He was clever at courting favour with influential and titled people and of charming the ladies. For example, the Duchess of Orléans said of him 'It's so rare for intellectuals to be smartly dressed, and not to smell, and to understand jokes.'[53]

Leibniz, like Descartes, is not well known for his contributions to physics. This is strange, as Leibniz spotted what Newton had missed in the '*crucis*' experiment of inelastic impact:

> [I maintain] that active forces are preserved in the world. The objection is that two soft or non-elastic bodies meeting together lose some of their force. I answer that this is not so. It is true that the wholes lose it in relation to their total movement, but the parts receive it, being internally agitated by the force of the meeting or shock…There is here no loss of forces, but the same thing happens as takes place when *big money is turned into small change*.[54]

The equivalence of this active force to our concept of energy is striking, especially when we realize that Leibniz's concept of force is not Newton's, but is given by the quantity mv^2. Leibniz called it '*vis viva*' or 'live force' (emphatically not the same thing as a 'life force'). Apart from the factor of ½, this is identical to our modern expression for kinetic energy. (To be fair to Newton, the latter had been referring to a collision between absolutely hard bodies, such as *atoms*, with no internal parts.)

But we must now go back and ask how Leibniz arrived at these ideas and what he meant by 'force'.

Leibniz was a rationalist, like Descartes, and built up a total picture of the world from considerations of how it ought to be rather than from just looking and seeing how it is. How 'it ought to be' meant being governed by overriding general principles, such as the Principle of Sufficient Reason, the Principle of Continuity, the Principle of Cause equals Effect and the Identity of the Indiscernables.

From the Principle of Continuity, there could be no finite hard parts—no atoms—and matter was continually divisible. It *was* therefore always possible to turn 'big money into small change'. Ultimately, every body had a non-material point-like origin or 'soul'—the monad. There was an infinity of such monads making up the universe and each was a centre of activity—a brain, if you like. Each monad followed its predetermined programme of action. Each also monitored its surroundings and kept pace with all the other monads, like perfect clocks all started together and so keeping perfectly in time with one another. In this way, the monads maintained the pre-established harmony set up by God. This had the strange consequence that the monads didn't need to (and in fact didn't) interact with each other:

> no created substance exerts a metaphysical action or influence upon another . . . All the future states of each thing follow from its own concept. What we call causes are in metaphysical rigour only concomitant requisites.[55]

This is strange indeed considering that the most influential precept of Leibniz's dynamics was to be that the total cause equals the total effect.

How did Leibniz's concept of monads lead to the observable features of the world? Descartes' gravity was caused by vortical pressures in the ether and Leibniz explained the observable properties of matter (hardness, elasticity, etc.) by an infinite regress of such vortices. Unlike Descartes, however, extension was not a fundamental property of matter (as, wearing his logician's hat, Leibniz insisted that 'extension' is a predicate that can only be ascribed to a class, and no substance is a class). 'Motion', also, was an apparent quality and not fundamental. The underlying reality was the infinite sea of souls . . . But what bearing does this have on physics and how did it lead to the opening quote on inelastic collisions, so rooted in the observable world and in 'common sense'?

The monads, as shown above, were centres of activity. This was the fundamental essence of matter in Leibniz's world. Matter was anything but passive:

> Substance is a being capable of action [and] to be is to act.[56]

(This, of course, was anathema to the mechanical philosophy.) Thus, 'activity' was the primitive agent and cause of all 'effect' in the universe. Clearly, from the Principle of 'cause equals effect', the total activity in the universe was conserved—but what was its measure? It could not be Descartes' quantity of motion (mv) as this had been shown not to be conserved in all cases. Once again, the testing ground was the case of impact between bodies. Whereas the *directed* 'quantity of motion' (our momentum) was known to be conserved in impact, Leibniz felt that the true absolute measure had to be a quantity without direction, as the Principle of Sufficient Reason would demand a reason for one particular direction to have been singled out. Also, something so fundamental as activity had to be always a positive quantity.

Searching for such a suitable absolute measure of 'force', Leibniz at last came upon it in the work of Huygens. (Leibniz frequently had discussions with his friend and contemporary, Huygens, at the Académie Royale des Sciences in Paris—Huygens was 17 years older than Leibniz). In Huygens' work on elastic impacts a certain quantity, total mv^2, was found to be conserved. This quantity, mv^2, had all the right specifications for Leibniz: it was conserved (in elastic collisions), it was a pure magnitude (had no direction), it stemmed from Galileo's results for freely falling bodies and, ultimately, it stemmed from the impossibility of perpetual motion—in other words, from the requirement that the cause should be equal to the effect. Leibniz had no hesitation in recognizing mv^2 as the absolute measure of active force that he had been looking for.

Leibniz, like Huygens and others, was inspired to correct Descartes. He gleefully sent off a paper entitled 'A brief demonstration of a famous error of Descartes and other learned men, concerning the claimed natural law according to which God always preserves the same quantity of motion; a law which they use incorrectly, even in mechanics'[57]—the title was anything but brief.

First, Leibniz managed to derive his mv^2 formula. (Descartes had found that mh = constant; merging this with Galileo's $v^2 \propto h$, Leibniz then obtained mv^2 = constant, where v was the speed when a body of mass m fell from rest through a height h.)

Then, Leibniz went on to show that Descartes' quantity of motion, mv, led to error. For example, the combinations ($m = 1$, $v = 4$) and ($m = 4$, $v = 1$) (arbitrary units) had the same quantity of motion, but in the first case the mass could ascend to a height of 16 units, whereas in the second case the mass could ascend to a height of only 1 unit.

These heights were different, and so Descartes was opening the door to 'perpetual motion' and a denial of 'cause equals effect'.

It was somewhat disingenuous of Leibniz to trumpet the fact that the cause didn't equal the effect, as if Descartes would have been unconcerned by this. Nevertheless, Leibniz had made a valid point concerning the contradiction between the two measures of force. Descartes was dead by this stage, but his followers (including Papin of the 'digester') tried to defend his position. They claimed that the time of fall should be taken into account. Leibniz countered that time wasn't relevant—the total effect would be the same whatever the time of fall or route taken:

> For a given force can produce a certain limited effect which it can never exceed whatever time be allowed to it. Let a spring uncoil itself suddenly or gradually, it will not raise a greater weight to the same height, nor the same weight to a greater height.[58]

However, time *was* very relevant to Leibniz's dynamics—it was the medium through which the effect was developed. Leibniz said that in the static situation (a balanced lever, a dead weight, etc.) there was a propensity to motion or *conatus* (Leibniz took the word from Hobbes). This propensity to motion was due to a force that Leibniz referred to as the dead force, or '*vis mortua*', and which he claimed had the same measure (mv) as Descartes' disputed quantity of motion:

> I call the infinitely small efforts or *conatus*, by which the body is so to speak solicited or invited to motion, solicitations.[59]

After the motion has started, the infinitesimal *conatuses* are compounded through time in a gradual and continuous way. Leibniz was able to use his own integral calculus to accomplish this:

> what I call dead force... has the same ratio in respect to living force (which is in the motion itself) as a point to a line.[60]

The *vis mortua* does bear some resemblance to our modern potential energy, but its measure (mv) is wrong. According to Westfall,[61] Leibniz rigged the formula so that *vis viva* would come out right (for readers with calculus, Leibniz defined *vis mortua* as mv so that the integral of it with respect to v would come out as mv^2; in fact, as $\frac{1}{2}mv^2$). Also, Leibniz showed no inkling of the continual interplay between the *vis mortua* and the *vis viva* whereby an increase in one is accompanied by a decrease in the other. In the case of free fall, for example, $v^2 \propto h$ was understood by Leibniz to apply only at the end of the motion—when the force was totally exhausted in ascent or fully developed in descent. More importantly, whereas our measure for potential energy is derived from an integration over space, Leibniz's was compounded in time.

So, first, Galileo had battled for years trying to find a relationship between the speed acquired in free fall and *distance* and then Leibniz saw the 'live force' as acting through *time*.

It was nonetheless a truly impressive achievement of Leibniz's to recognize the important role of an active principle and to identify this as the quantity mv^2. What to Huygens had been nothing more than a number that happened to come out as a constant took on, in Leibniz's hands, a truly cosmic significance. It was the measure of all activity, all actual motion whatever the source—falling due to gravity, the elasticity in impact, the restoring force of a spring, and so on.

However, the evidence in favour of mv^2 as an absolute measure of force was not so overwhelming as it now appears to us. For one thing, mv^2 was not absolutely conserved, but was still only a constant for a given viewpoint (frame of reference). To this extent, it was no more useful a measure of force than momentum. For another thing, mv^2 was derived by reference to Galileo's calculations from free-fall experiments near the surface of the Earth. Why should it apply to other kinds of activity, for example, horizontal motion, or celestial motions which were not Earth-bound? Johann Bernoulli[62] was to make just this objection (see Chapter 7), saying that God could have made gravity other than he did at the Earth's surface. Finally, in the vast majority of cases *in practice,* mv^2 was *not* conserved.

But it was in just these real-life cases of inelastic collisions, where 'big money is turned into small change', that Leibniz's physical intuition really came to the fore. He was the only one of his contemporaries who could truly isolate what was going on from the obscuring details of real life. Frictional effects, Leibniz saw, were *always* present and a 'perpetual motion' was 'without doubt contrary to the order of things.'[63] At the same time, the universe *must* be self-sustaining (so Newton's solution of a 'reformation' was not to be tolerated). Once again, the Principle of Cause equals Effect came to Leibniz's aid: whereas a perpetual *mechanical* motion could not be allowed (else the effect would be greater than the cause), a perpetual *physical* motion had to apply (else the effect would be less than the cause). Thus, Leibniz saw, there could only be an equality between cause and effect:

A perfectly free pendulum must rise to exactly the height from which it fell, otherwise the whole effect would not be equal to its total cause. But as it is impossible to exclude all accidental hindrances, that movement soon ceases in practice, otherwise that would be a mechanical perpetual motion.[64]

Apart from the fact that Descartes' rules for impact disagreed with experiment, Leibniz had philosophical objections as well. Descartes had impact between unequal bodies as a special case, but this case was not reached continuously from the case of impacts between identical bodies. In Huygens' analysis, also, Leibniz's Principle of Continuity ('nature makes no leaps'[65]) was compromised by instantaneous changes of direction. But in Leibniz's philosophy there were no perfectly hard bodies:

if bodies A and B collide . . . they are . . . gradually compressed like two inflated balls and approach each other more and more as the pressure is continually increased; but . . . the motion is weakened by this very fact . . . so that they then come entirely to rest. Then, as the elasticity of the bodies restores itself, they rebound from each other in a retrograde motion beginning from rest and increasing continuously, at last regaining the same velocity with which they had approached each other, but in the opposite direction.[66]

This is a completely modern outlook. More than just replacing hard bodies by deformable ones, Leibniz replaced Huygens' forceless kinematics by a new dynamic analysis. It was, in fact, Leibniz who coined the word 'dynamics'.

Overview

Galileo brought in 'time', the relativity of motion, and the relationship $v^2 \propto h$; Descartes brought in the measures 'quantity of motion' (mv) and 'work of a machine' (weight × height); Huygens discovered that mv^2 was a constant; his appeal to symmetries and conservation laws was to pay dividends in a later century. Newton brought 'force' into an anaemic forceless world and founded Newtonian Mechanics. Finally, Leibniz seized upon mv^2 as 'live force', the cause of all effect in the universe. His new dynamic analysis was modern in all respects except for one—he didn't recognize the 'motion of the small parts' as *heat*.

4

Heat in the Seventeenth Century

The Nature of Heat

From Plato (around 400 BC) through to Bacon, Descartes, Boyle, Hooke, and Newton in the seventeenth century, heat was understood as being due to the motion of the constituent 'small parts'. Leibniz, as we have seen, saw inelastic collisions as leading to the motion of the 'small parts'. But no one made any link between these two motions.

There were other problems that would need to be resolved before heat could be recognized as one of the blocks of energy. For example: What is the quantitative measure of heat? Is heat a substance? What is the connection between heat and temperature? As energy is a difficult and abstract concept, so each of its blocks will be difficult and abstract.

The first 'modern' to see heat as a form of motion was Francis Bacon (1561–1626). Bacon put forward the view that:

> heat itself, its essence and quiddity, is motion and nothing else.[1]

He went on to observe that 'pepper, mustard and wine are hot'. Bacon became famous for introducing the method of induction in science and stressing the need for experimentation. However, he never carried out any experiments himself except for the one where he is reputed to have tested the preservative effect of stuffing a chicken with snow. He caught a chill and died.

Descartes attributed heat to the motion of the small parts, yes, but not directly. The small parts were set in motion by the subtle ether, his 'first element'. This 'element of fire' may have been the very start of the heat-is-a-fluid theory.

Descartes' contemporary and fellow adherent of the mechanical philosophy was the French priest Pierre Gassendi (1592–1655). Gassendi had a material theory of cold (his 'frigorific' particles) as well as of heat.

He wasn't very consistent, as he went on to write that 'Anaximenes and Plutarch thought that neither hotness nor coldness should be thought of as occupying a substance, but rather the atoms themselves, to which must be attributed all motion and hence all action.'[2]

Gassendi's main legacy was in his promotion of the ancient Greek idea of atomism, put forward by Democritus and Epicurus and preserved in the poem 'On the nature of things' by the Roman poet Lucretius. This magnificent poem had lain dormant for around 1700 years until Gassendi rediscovered it. Atomism and its need for a void (between atoms) was abjured by the Scholastics and Gassendi had to recast it in a form which didn't bring on a charge of heresy. Epicureanism was so maligned at the time that it has never fully recovered and even today is associated merely with having a gourmet palate.

Robert Boyle (1627–91) is most famous to physicists for Boyle's Law—that the pressure of a gas is inversely proportional to its volume. We'll come to this later. Boyle's main passion was chemistry, and his aim was to free it from all alchemical sympathies and antipathies, and from Aristotelian forms and qualities. He was scion to a fabulously wealthy Anglo-Irish family, and had his own laboratory and assistants and no pressure to earn a living. A contemporary wit referred to him as the father of chemistry and son of the Earl of Cork.

Boyle was a champion of the mechanical philosophy where all effects were due to 'those two most catholic principles of matter and motion'.[3] Heat, considered as the motion of the small parts, was therefore a perfect example of an effect to be explained within the mechanical philosophy. But what exactly was the evidence for this motion? As Maxwell was to put it, some 200 years later:

> The evidence for a state of motion, the velocity of which must far surpass that of a railway train, existing in bodies which we can place under the strongest microscope, and in which we can detect nothing but the most perfect repose, must be of a very cogent nature before we can admit that heat is essentially motion.[4]

Boyle appreciated that the question needed answering:

> in a heated iron, the vehement agitation of the parts may be inferred from the motion and hissing noise it imparts to drops of water or spittle that fall upon it. For it makes them hiss and boil, and quickly forces their particles to quit the form of a liquor and fly into the air in the form of steams. And... fire, which is the hottest body we know, consists of parts so vehemently agitated, that they

perpetually and swiftly fly abroad in swarms, and dissipate or shatter all the combustible bodies they meet with in their way.

He also made the perceptive observations that if the motion was to count as heat, then it had to be 'very vehement', its 'determination [direction] . . . very various' and that it applied to particles 'so minute as to be singly insensible'.

This internal motion could be engendered from the outside by friction, percussion, or attrition:

The vulgar experiment of striking fire with a flint and steel, sufficiently declares what a heat, in a trice, may be produced in cold bodies by percussion or collision.

Most interesting of all, Boyle made a distinction between heat and bulk motion:

though air and water be moved never so vehemently as in high winds and cataracts; yet we are not to expect that they should be manifestly hot, because the vehemency belongs to the progressive motion of the whole body

and

if a somewhat large nail be driven by a hammer into a plank, or piece of wood, it will receive divers strokes on the head before it grow hot; but when it is driven to the head so that it can go no further, a few strokes will suffice to give it a considerable heat; for whilst . . . the nail enters further and further into the wood, the motion that is produced is chiefly progressive, and is of the whole nail tending one way; whereas, when that motion is stopped, then the impulse given by the stroke . . . must be spent in making a various vehement and intestine commotion of the parts among themselves.

This is the only precursor that I can find to Leibniz's conversion of 'large money into small change' (Chapter 3).

Boyle did on occasion admit the possibility of Descartes' subtle element of fire. For example, he attributed the gain in weight of a metal upon burning ('calcining') to the absorption of fire atoms and, when adding water to salt of tartar, he says: '[the agitation was] vehement enough to produce a sensible heat; especially if we admit the ingress and action of some subtile and ethereal matter, from which alone Monsieur Des Cartes ingeniously attempts to derive the incalescence'.

It is interesting to note that Boyle only admitted a subtle heat substance in chemical processes. In the eighteenth and nineteenth centuries we shall find that the caloric (material or substance) theory of heat was mostly adopted by chemists while a dynamic theory of heat was mostly taken up by physicists (the terms 'physicist' and 'chemist' weren't introduced until the nineteenth century).

Finally, Boyle carried out many 'Experiments touching Cold etc.'[5] He noted the very large forces resulting from freezing ('congelation') and he speculated on the possibility of frigorific particles, but could find no change in weight in a block of ice as it thawed in the balance pan.

Robert Hooke (1635–1703) was Boyle's laboratory assistant and subsequently a natural philosopher in his own right. From 1662, he was employed as curator of experiments at the Royal Society. He is known to physicists for Hooke's Law—probably the least of his scientific accomplishments—rather than for his masterful work *Micrographia*, a book with many detailed drawings of the world revealed by his microscope.

Hooke was one who had no doubts about the redundancy of frigorific particles and fire atoms. In sarcastic vein he wrote:

> we need not trouble ourselves to examine by what Prometheus the element of fire came to be fetched down from above the regions of air, in what cells or boxes it is kept, and what Epimetheus lets it go; nor to consider what it is that causes so great a conflux of the atomical particles of fire, which are said to fly to a flaming body like vultures or eagles to a putrefying carcase and there to make a very great pudder.[6]

Thermometry

How should heat be quantified? And what is the relation of heat to temperature? Does a teaspoon of boiling water have less heat in it than a bath full of tepid water? Does a given temperature interval correspond to a given change in heat, and is this true anywhere within the temperature scale and for any thermometric substance or property? Is there a common zero of heat and temperature when all motion of the small parts has ceased? What is the connection between temperature, heat, and the *sensation* of heat?

Galileo made the point that the sensation of heat didn't exist by itself, it was only there if someone was around to do the sensing: 'there's no tickle at the end of the feather'.[7] But this was a truism. To answer the last question of the previous paragraph first, temperature *is* a quantification of the sensation of hotness—at least in the range of temperatures in which humans can do any sensing. It is a curious fact that there is no one physical property that fits this task of mapping out the temperature.

This is to be contrasted with the measure of pressure, which does have definite units (of force divided by area). Temperature only achieves this status when an absolute scale of temperature is defined—by Kelvin in the nineteenth century (Chapter 16).

Sensitivity of the skin is an obvious but subjective property. For centuries it was thought that ponds and caves were colder in summer than in winter, presumably as fingers were colder in winter than in summer. Volume change is the physical property that most readily presents itself as a suitable measure of temperature. This still leaves open the question of which substance to use.

Galileo chose the expansion of air for his thermometer (strictly, thermoscope—that is, an instrument giving only a qualitative indication). A student of Galileo's tells of the following lecture demonstration:

> Galileo took a glass vessel about the size of a hen's egg, fitted to a tube the width of a straw and about two spans long; he heated the glass bulb in his hands and turned the glass upside down so that the tube dipped in water contained in another vessel; as soon as the ball cooled down the water rose in the tube to the height of a span above the level in the vessel. This instrument he used to investigate degrees of heat and cold.[8]

(Galileo makes no mention of the beautiful device with colourful glass 'teardrops' now sold as 'Galileo's thermometer'.)

Jean Rey, a French country doctor, was the first to use the expansion of a liquid (in around 1630). He employed the same glassware as in Galileo's thermoscope, but with the water and the air swapped over. In a letter to Father Mersenne he explains:

> [my thermoscope is] nothing more than a little round phial with a very long and slender neck. To use it I put it in the sun, and sometimes in the hand of a feverish patient, after first filling all but the neck with water.[9]

(Father Marin Mersenne (1588–1648) was a Minim Friar, and acted as a sort of distribution network for scientific information in the mid-seventeenth century.) Rey's was the first liquid-in-glass thermoscope, but all such instruments open to the air were unreliable (the 'atmosphere' was soon to be discovered—see the next section). The first sealed thermometer was invented by the Grand Duke of Tuscany, Ferdinand II, of the famous Medici family. (Galileo had named the moons of Jupiter the Medicean stars after Ferdinand's father, Cosimo de Medici.)

Ferdinand was one of the founders of the scientific society, the Accademia del Cimento, in Florence in 1657, and his thermometer came to be known as the Florentine thermometer. It was widely distributed to other newly formed scientific societies and royal families across Europe. Boyle first came across one in 1661: 'a small seal'd weather-glass newly brought by an ingenious traveller from Florence, where it seems some of the eminent virtuosi that enobled that fair city had got the start of us in reducing seal'd glasses into a convenient shape for thermoscopes'.[10]

The duke found that ponds and caves were, after all, colder in winter. Edme Mariotte (*c.* 1620–84), of the French Royal Academy of Sciences, found that the cellars of the Paris Observatory varied by no more than 0.3° C (modern units) between winter and summer. Monitoring the temperature of these cellars became something of an obsession for French scientists right up to the French Revolution. For example, the Royal Astronomer, Jean-Domenique Cassini (1748–1845), went up and down the 210 steps almost daily in 1788 and 1789 to take readings.

Although water and mercury were investigated, spirit-of-wine (alcohol) was the preferred thermometric liquid as its freezing point was lower. The Florentine academicians found that water had its minimum volume just *above* the freezing point—Hooke flatly refused to believe this. (Water is therefore not useful as a thermometric substance in this region: the relation of volume to temperature may not be linear, but at least it should be single-valued.) Degrees were marked with beads of glass or enamel fused on to the stem but there were, as yet, no fixed points. Even so, the Florentine thermometers were of unsurpassed accuracy for the times. Amazingly, some 50 thermometers found in a trunk in Florence in 1829 were still all in agreement.

Fixed points were needed and snow, melting butter, blood-heat, boiling water and, of course, deep cellars were variously suggested. How should degrees be defined? Hooke and Huygens both tried defining a standard volume (at a fixed point) and then a degree represented a given fractional increase in this volume. The more reliable method of two fixed points was suggested by two Italians, Honoré Fabri in 1669 and Bartolo in 1672.

The question still remained: What was the thermometer measuring? In 1680 Francesco Eschinardi, another Italian, had the idea of having the heating effect of one standard candle to measure out one degree and the heating effect of two standard candles for two degrees, and so on. Here, at last, was a direct link between heat and temperature.

Carlo Renaldini (1615–79), Professor at Padua and former '*Accademician*', had a similar idea. He had two reservoirs of water, one boiling and the other ice-cold, and a one-ounce dipper. (There were 12 ounces in a 'pound' or Troy weight.) A mixture of one ounce of boiling water and 11 ounces of cold water was defined as having a temperature of one degree, two ounces of boiling water and ten ounces of cold water as having a temperature of two degrees, and so on.

Renaldini's method made the implicit assumption of a linear relation between degrees of temperature and quantity of heat. However, there were many results which didn't tie in with this. For example, the Accademia del Cimento found that the same volumes of different fluids at the same temperature caused different amounts of ice to melt. It wasn't until Joseph Black, 100 years later, that such problems were sorted out and a real understanding and robust measure of the quantity of heat began to emerge (see Chapter 6).

The Vacuum, the Atmosphere, and the Concept of Pressure

Boyle made the first sealed 'weather glasses' in England and wrote:

> At the beginning I had difficulty to bring men to believe there would be rarefaction and condensation of a liquor hermetically seal'd up, because of the school doctrine touching the impossibility of a vacuum.[11]

The 'school doctrine' were the Scholastics or churchmen who rigidly followed the teachings of Aristotle. They didn't believe in the vacuum: 'nature abhors a vacuum'. However, the mechanical philosophy required a vacuum between the atoms and also avoided emotive agents such as 'abhorrence'.

One bit of phenomenology that needed explaining was the long-known fact that syphons couldn't lift water over hills greater than 34 feet high and pumps couldn't raise water to more than this height. Galileo argued that the strength of materials was due to the 'force-of-the-void' binding atoms together and that the 'breaking strength'[12] of water was equivalent to a column of water 34 feet high (even the great Galileo had a blind spot[13]).

Evangelista Torricelli (1608–47), a brilliant young admirer of Galileo, joined the debate. He argued that it was the weight of the air rather than Galileo's vacuum-force that held up the column of water. As a simple balance of weights was involved, then the air should support the same weight of any liquid, but the height of the column would depend on the liquid's density. Torricelli experimented with seawater, honey, and finally mercury (this was mined in Tuscany). He ordered a glass tube, sealed at one end, and with extra-thick walls. He filled it to the brim with mercury, closed off the end with a finger, and inverted it into an open dish of mercury—the first barometer. The level of mercury in the tube fell until its height was 29 inches, or exactly 1/14th of the height of the column of water. As mercury was 14 times as dense, this result supported Torricelli's explanation.

However, it was at least 20 years before Torricelli's explanation was accepted. Some of the objections that he had to answer included the following:

(1) If the dish of mercury is covered over with a lid, why isn't the weight of the air cut off?
(2) Weight acts downwards, so how does the weight of the air go round corners and push the column of mercury upwards?

Torricelli had devised and played with hydrostatic toys—such as the so-called 'Cartesian diver'—and had built up an intuitive understanding of hydrostatic pressure. He not only answered the objections satisfactorily but went on to realize that the concept of pressure could apply to the air as well. He was therefore the first to appreciate the wondrous idea that 'we live submerged at the bottom of an ocean of air'.[14]

There was still the pertinent question of what was in the space above the mercury in the barometer (Fig. 4.1). (Torricelli, by the way, barely mentioned his barometer again—he had visited Galileo during the latter's house arrest and presumably wanted to avoid a similar fate.) The 'schoolmen' argued variously that the space contained a bubble of distended air, ethereal vapours emitted by the mercury or seeping in through fine pores in the glass—but why did just so much ether go through and no more?

Torricelli arranged for a glass tube to be blown with a large spherical shape at the sealed end. He found that whatever the shape and size of the space above the mercury and whatever the inclination of the tube, the mercury column remained at exactly 29 inches. What was more, by inclining the tube the space could be made to disappear altogether. In other words, the amount and shape of the 'ethereal vapours' made no difference—they just weren't there.

Blaise Pascal (1623–62), of Triangle fame, carried out a famous demonstration in the 1660s. He repeated the Torricellian 'balance of weights'

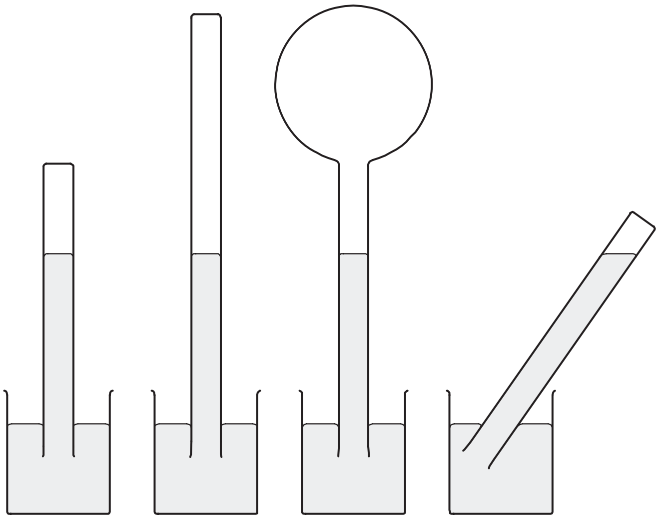

Fig. 4.1 Different barometers, variously tilted (redrawn from Westfall, *The Construction of Modern Science*, Cambridge University Press, 1977).

experiment but on a grand scale, using two glass tubes approximately 50 feet in length and supported against the mast of a ship in the harbour at Rouen (the glass industry in Rouen must have been very advanced). He filled one tube with wine, the other with water, keeping the air out all the while—one can only wonder at the practicalities of doing this. The crowd was asked to predict the result. The established view was that the wine, being more 'spirituous', should stand at a lower height than the water as the 'spirits' pushed the wine down. The 'vacuists', on the other hand, argued that the wine, being less dense, should stand at a taller height than the water. The result was that the water stood lower than the wine, as Pascal had expected.

Pascal concluded that nature did abhor a vacuum, but just to the extent of the weight of a column of mercury 29 inches high. After more experiments, however, such as the famous one where Pascal's brother-in-law took a barometer up to the top of the Puy de Dome in 1648, and the brilliantly ingenious 'vacuum within a vacuum', Pascal did concede that the vacuum exists. (The latter experiment clinched the irrelevance of spirituous ethers.)

Otto von Guericke (1602–86), from Magdeburg in Germany, constructed a sealed container made from two hemispheres and then used a water pump to pump the air out (this was thus the first *air* pump). He then demonstrated the force of the vacuum by the number of horses required to separate the halves. (Apparently, two teams of eight horses were required, so either the horses were exceptionally feeble or von Guericke was exaggerating the strength of his vacuum.)

Boyle joined in the debate and in his private laboratory in Oxford he asked his assistant, Hooke, to improve on von Guericke's air pump. Hooke did this and also made a 'receiver'—a bell jar that surrounded the exit-tube from the pump and rested on a plate. A special stopper in the bell jar allowed experiments to be introduced without destroying the vacuum. In the 1650s Boyle and Hooke carried out many such 'pneumatics' experiments, with the receiver variously evacuated, pumped with extra air, or with the same air inside as outside.

From experiments such as introducing a barometer or a carp's bladder into the receiver, Boyle reaffirmed that air had weight and that the vacuum exists. He also found out something new—the air had 'spring'.[15] For example, the carp's bladder slowly expanded when placed in a vacuum, presumably due to the expansion of residual air within it. Boyle concluded that air was an 'elastic fluid', which always expanded to fill any container.

The Scholastics, in the person of the Jesuit Father Linus, thought that they now had Boyle cornered: if air was supposedly elastic and expansible, it should also be compressible—plainly absurd, thought Linus. Boyle countered by carrying out the experiments for which he is known to posterity. A glass tube was curved into the shape of a J, with the shorter end being sealed. Some air was trapped in this shorter end by a column of mercury poured into the longer end. By increasing the amount of mercury, the 'spring' (pressure) on the trapped air could be increased and the volume change noted. The air was indeed compressed. The volume of the air for successive increments in the column of mercury supported 'the hypothesis that supposes the pressures and expansions to be in a reciprocal proportion'.[16] (Boyle's Law). (In France this is known as Mariotte's Law as Mariotte carried out similar experiments—but after Boyle.)

Thus, the vacuum had arrived, as had the atmosphere and the concept of the pressure of an elastic fluid. Pressure would turn out to be a concept almost—but not quite—as intimately linked to heat and to energy as was temperature.

Pneumatics and Further Studies on Heat

The air pump ushered in a host of 'pneumatic' experiments. The vacuum was found not to support life. Philosophers and invited guests watched with morbid fascination as various animals (a bird, a mouse, an eel) were put in the receiver and seen to die as the air was pumped out. The flame of a candle also went out in the vacuum and the sound of a bell became very faint, but magnetic and electrical effects persisted. Von Guericke, Boyle, and Johann Bernoulli noted an occasional luminous glow in an evacuated receiver but could not account for it. Heat was understood to be required for water to boil, but 'ebullition' was observed in a dish of merely tepid water as the air was pumped out.

Light was transmitted through the vacuum, enabling all these experiments to be witnessed (according to the Scholastics, if a vacuum could be made to exist at all, then it should be opaque). A 'burning glass' (lens) outside the receiver could be used to focus light on to something within the receiver and set it alight. In a celebrated demonstration, Mariotte caused gunpowder to explode using an ice lens. In this and other experiments it was shown that heat rays could be focused, gunpowder needed no air in order to burn, and that heat 'rays' as well as light could be transmitted through the vacuum. (This last result was hard to explain with the motion theory of heat.)

Mariotte, in his '*Essai du chaud et du froid*'[17] in 1679, describes other experiments showing the properties of radiant heat (as we would now call it). Dark colours and rough surfaces absorbed heat more readily and light-coloured shiny surfaces reflected the heat. Mariotte put a concave metal mirror near a fire and found that he could not long endure the heat when his hand was at the focus. However, when a glass plate was placed between the fire and the mirror, the heat could no longer be felt but the light of the fire was undiminished. This was the first time that the light and the heat from a given source had been separated. But when the Sun was the source, then the light and heat could not be separated.

We can begin to appreciate the enormous difficulties in identifying and separating out the various phenomena all associated with the one concept, heat. Before us moderns become too complacent—we do after all know how the story will turn out—consider the puzzling case of the reflection of cold. In the *Saggi di naturali esperienze* of the Accademia del Cimento, there is a description of an experiment whether a concave mirror exposed to a mass of 500 lbs of ice would make any perceptible reverberations of cold in a very sensitive 400-degree thermometer located at its focus. In truth this at once began to descend, but because the ice was nearby it remained doubtful whether direct cold or that which was reflected cooled it most. The latter was taken away by covering the mirror, and (whatever the cause may have been) it is certain that the spirit immediately began to rise again.[18]

With the hindsight of the Second Law of Thermodynamics, we can argue that the reflection of cold is impossible (see Chapter 16)—well, almost (see Chapters 17 and 18).

Edmond Halley (1656–1742) understood that the heating effect of the Sun's rays was proportional to the sine of the angle of incidence. He further understood that this led to a differential heating of the Earth's surface which then led to the trade winds. (Martin Lister, on the other hand, thought that the trade winds were due to the diurnal respiration of the Sargasso seaweed.) While Halley explained the winds as being due to hot airs being 'rarefied'[19] and therefore rising, he could not understand the process of evaporation: how could a heavy substance like water rise through a lighter substance like air? Halley suggested that the water atoms became distended and filled with a 'finer air'. Halley, Mariotte, and Claude Perrault (1613–88) all made estimates of the total rainfall, of the rate of flow of rivers, and of the evaporation to be expected from lakes and seas. In this way, it began to be appreciated that there was a closed cycle of water—so occult processes involving eternal springs and subterranean caverns were not needed.

While the vacuum did not support life or flame, increasing the air pressure within the receiver appeared to make animals more vigorous, plants grow more strongly, and burning coals glow more brightly. Rey found that gently heating lead and tin in air led to an increase in their weight: he proposed that some of the air had 'condensed'. Mersenne objected that, on the contrary, air was 'rarefied' upon heating.

Another doctor, John Mayow (1640–79), performed experiments that brought out common aspects of respiration and combustion: when a mouse in an evacuated receiver died, then a candle could no longer be ignited. However, the idea that some component of the air was being used up did not readily suggest itself to contemporary investigators. For one thing, the reduction in the volume of air was negligible. When Mayow placed a bell jar over water, then a burning candle caused the water level to rise by only 1/14th. To Mayow, this small reduction indicated a loss in the elasticity of the air. (In modern terms, about as much

CO_2 is put out as O_2 is taken up, so the initial difference in the volumes is very slight. However, the CO_2 is more soluble, it slowly dissolves in the water, and this eventually leads to an imbalance between the initial and final gas volumes.)

Mayow (and also Hooke) postulated that both respiration and burning required a nitrous spirit and that this could be found in the air. This 'nitro-aerial' spirit was present in a concentrated form in gunpowder and explained why the latter could burn without air. The nitro-aerial spirit was thought to be responsible in a more general sense for animal heat and motion and plant growth: a 'vital or life force' not to be confused—yet!—with Leibniz's 'live force'.

Heat-engines

The study of pneumatics and heat opened the way for a new technological advance—the heat-engine. The main incentive for the new 'fire-engine' (as it was usually called) was the need to pump water out of increasingly deep mines.

The Marquis of Worcester (we have heard of him being saved from the Tower of London so that he could perfect his perpetually rotating wheel; see Chapter 2) invented his 'water-commanding engine' in around 1663 (Fig. 4.2). Water was heated in a boiler and the steam collected in a chamber. When the heat supply was withdrawn, the steam condensed and water was sucked up from below. This water was then forced up a vertical pipe by the pressure of the next round of steam. The engine was erected at Vauxhall in London and raised water to a height of 40 feet. The Marquis obtained the rights to 99 years of use by Act of Parliament, but the engine was never used commercially (or at all?).

Of more success, in 1698, was the engine invented by Thomas Savery (1650–1715). This engine worked on the same principle as the 'water-commanding engine' except that externally applied cold water was used to speed up the condensation of the steam. A partial vacuum was created and then atmospheric pressure forced water into the engine-chamber and fresh steam drove this water up the up-pipe. In this way, mines could be 'drayned'—or so it was claimed.

Steam pressure was used directly by the French inventor Denis Papin (1647–1712) in his 'digester'—basically, a pressure cooker. He describes how beef bones could be softened by cooking in this manner:

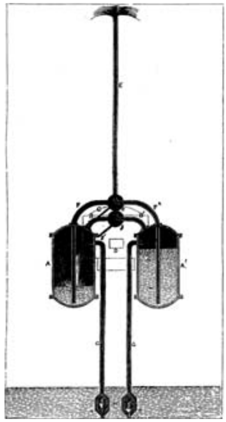

Fig. 4.2 The Marquis of Worcester's 'water-commanding engine', from *The Century of Inventions*, 1655.

I took beef bones that had never been boiled but kept dry a long time, and of the hardest part of the leg; . . . Having pressed the fire until the drop of water would dry away in 3 seconds and ten pressures, I took off the fire, and the vessels being cooled, I found very good jelly in both my pots . . . having seasoned it with sugar and juice of lemon, I did eat it with as much pleasure, and found it as stomachical, as if it had been jelly of hartshorn.[20]

An important development was the transfer of the piston-in-a-cylinder technology of the air pump into the heat-engine. This happened via the following route. First, von Guericke suggested the transmission of power from a piston in a long, narrow cylinder connected to one in a short, wide cylinder (the same principle of operation as used in today's air brakes). Then Huygens imagined a piston in a cylindrical chamber for his hypothetical gunpowder engine. Papin then conceived a method of transmitting power over long distances, using compressed air travelling through a tube (something like the system of canisters in pipes employed in the department stores of 50 years ago?). Finally, in 1690, Papin proposed an engine based on Savery's design but with steam condensing in a cylinder and drawing down a piston. (He was by this time a Huguenot exile in London, employed as curator of experiments at the Royal Society at the invitation of Huygens.) Papin was thus the very first proposer of the cylinder steam engine. It is probable that Thomas Newcomen, of the famous Newcomen engine, had never heard of Papin's proposal when he came to invent his own cylinder steam engine in 1712 (for more details, see Chapter 5).

In the final year of the seventeenth century, a different sort of engine was put forward—by Guillaume Amontons (1663–1705).

Amontons, son of a lawyer from Paris, had a childhood illness that left him almost completely deaf. Despite some opposition from his family, he devoted himself to mathematics and physical science. Perhaps his deafness was the inspiration behind his idea of a system of hilltop signalling using spyglasses and separate signals for each letter of the alphabet. It's not certain whether this was ever carried out (there was reputedly a demonstration in front of Louis XIV), but it would have been a first for telegraphy.

A definite first was Amontons' invention of a heat-engine using the expansion of air to provide the motive force.

Amontons had already been impressed by the large expansivity of air and had exploited this in his air thermometer. This was like a barometer except that it was sealed at *both* ends. A pocket of air at one end was warmed and its temperature increased: its pressure also increased and this raised a column of mercury. It wasn't possible to ensure that the air's volume remained exactly constant, but Amontons nevertheless managed to establish a rough proportionality between temperature and pressure for air between the ice and boiling points of water (this later became known as Charles' Law). Amontons also made careful investigations of the amount and rate of expansion of air on heating, and surmised that if the pressure could be kept constant then the volume would increase in proportion with the temperature.

Amontons noted that if expanding air could move a column of mercury, could it not also be made to raise water for a machine? This led him on to considering 'fire-engines'. His design of a 'fire-mill', as he called it, consisted of a wheel with a hollow rim and spokes, divided into cells with interconnecting valves. The wheel was filled with water on one side and air on the other. A fire was lit on the air side and the heated air expanded and forced water through the cells. This caused the wheel to overbalance and start to rotate. (The design possibly drew from his adolescent interest in perpetual motion machines.)

Amontons had grand plans to construct a 'fire-mill' some 30 feet in diameter. This was never realized although he did make a small-scale model. But Amontons went further than any of his contemporaries in appreciating the performance criteria of a generalized heat-engine. He commented on the tremendous power of heat, such as in the 'violent action' of the cannon:

> Nobody doubts that fire gives rise to violent action... But could it not also be used... to move machinery where one usually employs limited animate forces, such as men or horses?[21]

More than this, Amontons made an attempt at quantification. From his experiments on the heating and cooling rates of air and water, he estimated that his wheel would rotate once in 36 seconds and that it would exert a force sufficient to replace 39 horses. By observing glass polishers and their rate of working, Amontons estimated that the power of a man is 'about a sixth part of the labour of a horse'. By 'labour of a horse' he meant the 'force exerted by the horse multiplied by the speed of the horse at its point of application'.[22] This is essentially the same as our modern definition of power, or the rate of doing *work*.

Amontons also estimated what would later be called the 'duty' of the engine—the work done for a given amount of fuel—and claimed that his fire-mill would compare favourably with the pre-existing sources of power; namely, wind, water, animal, and man.

Finally, Amontons tackled for the first time the problem of losses in duty because of friction. He carried out experiments and found that:

(1) Friction did not depend on the area of the surfaces in contact, but only on the weight on those surfaces.
(2) For lubricated surfaces, friction was roughly independent of the materials.
(3) For frictional losses in a machine, the important thing to calculate was the 'couple' (weight × distance) of the frictional forces at the bearings. So, in hauling up a weight over a pulley, the actual power loss would be around 1/30th of the total power (he estimated). Amontons also estimated that in the case of his own fire-mill the frictional losses would be of the order of 1/8th.

These thoughtful experiments and insights surely earn Amontons the title of the grandfather of thermodynamics.

Overview

The change in outlook in European science in the seventeenth century is sometimes referred to as the Scientific Revolution. We have already seen (Chapter 3) the birth of mechanics culminating in Newton's great work, the *Principia*. The air pump, telescope, and microscope were invented during this century, and all opened up new worlds. Also, the new clock, thermometer, and barometer helped to chart the parameters of this revolution. In particular, the last two instruments introduced the new pneumatic variables of pressure, P, volume, V, and temperature, T, which describe the properties of an elastic fluid. The relations $P \propto 1/V$ (Boyle's or Mariotte's Law), $P \propto T$ and $V \propto T$ were found to hold, more or less.

The phenomenology of heat had increased enormously. It was understood to have a role in expansion, evaporation, fusion, and boiling; and was an agent in chemical decomposition, as in combustion, fermentation, and putrefaction. Heat also had links with light and with life. In fact, heat had so many facets that it was impossible to embrace or even recognize all the phenomena within one concept and one body of theory. But all were agreed that heat is the motion of the 'small parts'. This consensus would still leave room for two competing theories to emerge in the eighteenth century (see Chapter 8).

While the motion of the small parts was taken for granted it was still not agreed what the fundamental measure of this motion was, mv or mv^2 (this controversy is covered in Chapter 7, Part I). The question of whether these macroscopic measures could apply to microscopic parts as well as to everyday bodies had not been asked or even dreamed of (except, perhaps, by Newton: in his *Opticks*, Query 31, he speculated that all microscopic effects might be explainable within his Laws of Motion).

In the very last year of the century there were three designs of heat-engine—using atmospheric pressure, steam pressure, and the expansion of air. While none was in the least bit practicable (except perhaps for Papin's digester), there was an understanding by a few visionaries that engines were the way of the future. Amontons, in particular, envisaged a generalized heat-engine that could do 'work', and that had given fuel requirements and a given efficiency. No one, of course, had any idea of the role that the heat-engine would have in the understanding of energy.

5

Heat in the Eighteenth Century

Hot Air

In the eighteenth century, heat and gas studies were inextricably bound together. The two main theories of heat hinged on different conceptions of what a gas was. The mechanical (later dynamical, motion or kinetic) theory of heat saw heat as a process—the motion of the constituent particles—and its first quantitative working out was in the gaseous case. On the other hand, the substance or material theory of heat had heat as a thing, a subtle elastic fluid or ether—in other words, a gas of sorts.

The phenomenological links between heat and gas were becoming more and more evident. Combustion and also 'calcination' (in modern terms, oxidation) were known to occur only when certain 'airs' were present. Also, various 'airs' were released on heating, fermentation, and rotting. Respiration was necessary to life and was somehow connected with animal heat. Heat was connected to the processes of evaporation and boiling. Perhaps the most telling feature linking heat and gas was the enormous expansivity of water as it turned into steam and of air when heated—both exploited in various sorts of heat-engine. The fact that 'fire' and 'air', two of Aristotle's four elements, shared this property of being very elastic was thought to be highly significant by most philosophers at this time.

Guillaume Amontons (see Chapter 4) had shown that air increases its volume on heating by around one-third between 0°C and 100°C (modern units) and therefore, assuming linearity, by around 1/300th per degree. It seemed as if, at last, the perfect thermometric substance—better even than mercury—had been found. We now understand that the reason why gas has this special link with heat is that the forces between gas particles (molecules) are pretty well negligible (at usual densities) and so we see 'naked heat'.

Our modern view of a gas is roughly as follows: a bucket of gas at room temperature and pressure contains around 10^{23} molecules weighing around 10^{-26} kg each, moving in random directions with speeds of a few thousand kilometres per hour.

The eighteenth-century conception of a gas was utterly different to this. The particles of the 'aeriform fluid', as it was often called, were essentially static. Each particle occupied a definite location and had a tiny jiggling motion about this mean position (I call this the 'passion-fruit jelly' model). Even well into the nineteenth century, James Dalton (of atomic theory fame) had a completely static view of a gas. But what then could account for Boyle's 'spring of the air'?

Boyle had shown (see Chapter 4) that the air had 'spring'—it expanded to fill its container. He says:

> [I conceive] the air near the earth to be such a heap of little bodies . . . as may be resembled to a fleece of wool. For this . . . consists of many slender and flexible hairs; each of which may indeed, like a little spring, be easily bent or rolled up, . . . still endeavouring to stretch itself out again.[1]

But Newton realized that Boyle's ramous particles would only account for the air's resistance to compression and some small springiness, but would not account for its tendency to expand without limit.

Newton conjectured that all phenomena were due to just two kinds of forces—forces of attraction or of repulsion. The expansive property of air might be due to a repulsive force between the air particles and (Newton calculated) if this repulsive force acted only between nearest neighbours and varied in inverse proportion to the distance separating these neighbours, then the air would satisfy Boyle's Law. However, Newton admitted that he didn't know if the air really *did* behave like this:

> But whether elastic fluids do really consist of particles so repelling each other, is a physical question . . .[2]

Newton also toyed with the idea of ethers (he spells it 'aether'), for example, an ether that could transmit gravity without the need for action-at-a-distance. Such was the authority of Newton that despite his circumspection his followers embraced his theories and forgot their speculative origins. His model for a gas and his 'aether' gave legitimacy to a subtle-fluid theory for heat. The expansive property of heat could then be explained by the repulsive force acting between neighbouring 'fire particles'. Newton himself still preferred a mechanical theory of heat (as evidenced, for example, in his mechanical explanation of evaporation).

Newton's insights often sound so crisp and modern (especially compared with Boyle's longwindedness), such as when he touches on black bodies in the *Opticks*. He suggests that they heat up more than bodies of lighter colour because the incident light is 'stifled and lost'[3] within them.

Newton's most famous contribution to heat studies is his Law of Cooling[4] despite the fact that he published it anonymously (he had a reputation to uphold and was, in any case, a very guarded and secretive person). The law claimed that the rate of cooling of a body was proportional to its excess temperature over its surroundings. Newton's purpose was to use this law to extend the temperature scale to regions above the capability of mercury-in-glass thermometers: the times taken for various molten metals to cool down and solidify gave fixed points on a line that extended to very high temperatures.

In the year in which Newton died (1727) Hales published his *Vegetable Staticks*.[5] He was the first investigator to give priority to the study of gases. (One might say that in going from Thales to Hales one goes from 'water is all' to 'gases are all'.) In fact, Hales always used the term 'air' or 'airs', even though Jan Baptist van Helmont in the sixteenth century had already coined the word 'gas', taking it from the Greek root *kaos*.[6]

Stephen Hales (1677–1761) was a physician and clergyman from Teddington, Middlesex. (The poet Alexander Pope, from neighbouring Twickenham, makes mention of 'plain parson Hales'.)

Hales heated many substances, including hog's blood, amber, oyster shells, beeswax, wheat, tobacco, gallstones, and urinary calculi, and collected the yields of 'air'. He invented the pneumatic trough for this purpose (Fig. 5.1). The trough used water, which resulted in some soluble gases being lost. However, Hales had no idea that there were *different* gases with different chemical properties: he saw the purpose of the water merely as a method of washing the air clean of impurities. (Later in the century, Cavendish and Priestley would improve on the trough by collecting gases over mercury.)

Hales made the startling discovery that prodigious quantities of gas were released when things were heated. For example, an apple yielded 48 times its volume of 'air'. Hales commented that if the air had

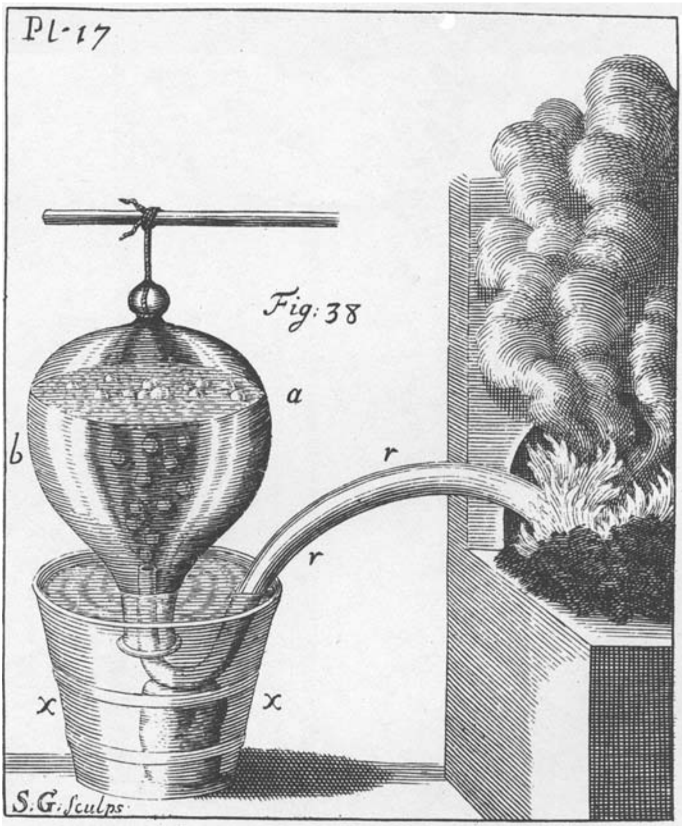

Fig. 5.1 Hales' pneumatic trough, adapted from *Vegetable Staticks*, 1727 (republished by the Scientific Book Guild, Oldbourne Book Co., 1961).

to be forced back inside, presumably requiring a pressure of some 48 atmospheres, then the apple would explode. He concluded that there must be a repulsive force in 'elastick air', but this was converted into an attractive force when the gas became 'fixed' within the solid—air was 'amphibious',[7] he said.

Hales made no attempt to identify the different airs. In the early eighteenth century air was still understood to be variable in quality—like soil, say. (Water, on the other hand, was taken to be elemental.)

The Phlogiston Theory

Combustion was another arena where heat and gas studies overlapped and where the concept of an ether was again invoked. It was noticed by Boyle, Hooke, Mayow, and others that combustion could only occur in air (apart from special cases such as the combustion of gunpowder; see Chapter 4). It was also noticed that the volume of the air was slightly reduced after combustion had taken place. It may seem obvious to us now that a component of the air is being used up, but this was not the natural inference to make in the eighteenth century (for one thing, oxygen had still to be discovered). The contemporary explanation was a physical rather than a chemical one—that the elasticity of the air had lessened.

It was nevertheless thought that a vital ingredient was required for combustion (Mayow called this the nitro-aerial spirit). It was supposedly present in air (and also in gunpowder) and was responsible for the air's elasticity. An alternative view developed in Germany whereby the combustible essence was taken to be *in* the burning body rather than in the air. Everything about this alternative combustion theory was an inversion of the former approach.

First, Becher in 1667 put forward his theory that combustible bodies contain a substance ('oily earth') that gets used up and released during combustion, leaving a stony 'vitreous earth' residue. Then Georg Ernst Stahl, in 1703, revived Becher's theory and called this combustible essence 'phlogiston' after the Greek prefix *phlog*, or *phlox*, for flame. Charcoal was thought to be an almost pure embodiment of phlogiston. Part of the evidence was that when charcoal burns it releases the phlogiston and only a few ashes are left behind.

When a metal is heated in air then a 'calx' is formed—what we now know to be the metal oxide. When the calx is heated in the presence of charcoal, the phlogiston from the charcoal is absorbed by the calx and combined with the metal. The phlogistonists saw the fact that the metal was shiny, lustrous, and malleable as evidence that it contained more phlogiston ('fiery essence') than the dry, brittle calx.

One problem presented itself straight away: the calxes weighed more than their respective metals (Dr Rey and Boyle had noticed this) even though the metals were now *combined with* phlogiston. The solution was to give phlogiston the property of negative weight—or 'levity'.

The Material Theory of Heat

One who did not concur with the phlogiston theory was Hermann Boerhaave (1688–1738) of Leyden in Holland, but Boerhaave was the initiator of another subtle-fluid theory—the material theory of heat. His book, *Elementiae chemicae*[8] (1735), was translated into French, German, and English, and was widely distributed and read. In it he established the morphology of heat studies for the next 30 years.

Boerhaave was a remarkable and influential person in other ways as well. In medicine, he was the founder of a great teaching tradition at Leyden and he was the first person to instigate a systematic approach to clinical practice (he introduced the use of the Hippocratic Oath, the post mortem, and Fahrenheit's thermometers). His capacity for hard work was phenomenal. He held three professorships simultaneously at the University of Leyden—in medicine, chemistry, and botany—and showed almost fanatical zeal in his experiments on mercury, undertaken to show that the transmutation of mercury into gold was not possible. (The mercury was purified by straining it through leather, washed in seawater and then shaken in a fulling mill for 8½ months. In another experiment, a gold-mercury amalgam was distilled 877 times, and in yet another some mercury was heated continuously for 15½ years—no changes in its weight or purity were detected.)

In the *Elementiae chemicae* Boerhaave made no reference to his contemporary, Stahl, or to Stahl's phlogiston theory, and he rejected Stahl's whole ideology of animism. On the other hand, Boerhaave very much admired the English philosophers, Boyle and Newton, and their adherence to the mechanical philosophy. But Boerhaave still didn't go along with their mostly mechanical view of heat. In this regard he looked even further back, to Descartes' element of fire, and Boerhaave can therefore be regarded as the founder of the material, or fluid, theory of heat.

Boerhaave attributed heat—or 'fire' as he called it—with truly cosmic significance:

> Fire is the universal cause of all changes in nature... Thus were a man entirely destitute of heat he would immediately freeze into a statue.[9]

The fire particles were too minute to be visible and were weightless. Boerhaave checked up on this by weighing hot iron continuously as it cooled and was re-heated. There was no detectable change in weight.

For Boerhaave, it was axiomatic that 'fire' could not be created or destroyed:

> The quantity of fire in the universe is fixed and immutable.

He therefore did not agree with Boyle that 'fire' could be created anew by percussion, friction, or attrition: 'Fire…is a body *sui generis*, not creatable, or producible *de novo*'. But Boerhaave could not deny the common experience of heating in such cases and he gave the rather lame (or, at least, odd) explanation that attrition is very swift, and only fire—the swiftest thing in nature—can keep up.

A controversial aspect of Boerhaave's views was his 'distribution theory' of fire—that fire is to be found equally everywhere:

> Pure or elementary fire is equally present in all places; nor is there any point of space, or any body, wherein there is more found than in any other.

Even the coldest regions would therefore contain fire:

> For if, even in Iceland, in the middle of winter, and at midnight, a steel and flint be but struck against each other… [then] a spark of fire will fly off.

He agreed that an absolute zero of temperature would imply a total absence of heat, but thought that this could never happen in practice. Another thing that Boerhaave deemed to be against nature was that different substances should attract heat to differing extents:

> Among all the Bodies of the universe that have hitherto been discovered and examined, there never was yet found any one that had spontaneously, and of its own nature, a greater degree of Heat than any other.
>
> And this gives us occasion to adore the infinite wisdom of the great creator. For had there been anything that attracted fire; it would have become a perpetual furnace: and how unhappy must the condition of man have been under such circumstances?

Higher and lower temperatures could still occur in forced cases, and they implied, respectively, higher or lower densities of the heat-fluid.

Boerhaave understood the difficulty of defining the quantity of heat and noted that 'altho' we are able to discuss the Force of Fire by its sensible effects; yet we cannot from its Force make certain judgements of its quantity'. Nevertheless, he did attempt to measure these sensible effects, the chief of which was the expansion or contraction of the given 'thermometric' substance. He asked Fahrenheit to perform the experiments.

Daniel Gabriel Fahrenheit (1686–1736), famous for his temperature scale, didn't have a very auspicious start to his career. His parents died of mushroom poisoning and his legal guardians arranged for him to train as a book-keeper; but Daniel wasn't interested, stole some money, and ended up with a warrant being issued for his arrest. His guardians attempted to pack Daniel off to the East Indies, but Daniel managed to evade both arrest and the East Indies and subsequently led a peripatetic existence between Germany, Sweden, and Denmark, making thermometers all the while. He eventually settled in Amsterdam (in around 1717), the warrant having been cancelled by this time.

Fahrenheit carried out various experiments for Boerhaave using, on some occasions, the drying rooms of the local sugar refinery as a source of heat. For example, a dog, cat, and sparrow were observed as they died of heat exhaustion. On a less gruesome note, Fahrenheit examined the method of mixtures which had been initiated by Renaldini in Padua (see Chapter 4) and investigated further by Brook Taylor (1685–1731) in London. Taylor mixed given quantities of boiling and freezing water and found that the temperature of the mixture was the arithmetical mean. (Brook Taylor was the inventor of the Taylor series, an idea that sparked from discussions in a London coffee house,[10] and was a mathematician and gentleman-scientist of private means.)

Fahrenheit tried mixing mercury and water together, but the final temperature was nowhere near the mean—in fact, it was always closer to the water temperature. By trial and error, Fahrenheit eventually found out what proportions, by volume, of water and mercury were required in order to make the final temperature equal to the mean. He said in a letter to Boerhaave:

> I changed the proportions, taking instead [of] equal volumes, three volumes of quicksilver and two volumes of water, and then I obtained roughly the mean temperature…A most remarkable aspect of this experiment is that although the quantity of matter in the quicksilver in relation to the water is about 20 to 1, its effect is no greater than if water were mixed in the proportion of 1 to 1. I leave the explanation of this to you, Sir, and would only add that it may be possible to deduce herefrom the reason why quicksilver thermometers are more sensitive than alcohol thermometers.[11]

In other words, despite the fact that the quicksilver was some 20 times denser than the water (the correct figure is 14 times denser), a temperature halfway between the two initial temperatures was only obtained when water and mercury were mixed in the volume ratio of 2 to 3.

Boerhaave conceded that 'the matter of fire' was clearly not distributed in proportion to density, although this had been his first guess. However, he took the ratio of 3 to 2 as sufficiently close to unity that he was prepared to conclude that 'fire' *is* distributed in direct proportion to volume—his 'distribution theory of heat'.

George Martine (1700–41), a young Scottish physician from St Andrews, was not convinced by Boerhaave's conclusions. He knew Boerhaave and Boerhaave's work very well as he had obtained his doctoral degree from the University of Leyden. However, Martine's approach was less influenced by metaphysical speculations. He always used the pragmatic term 'heat' rather than the emotive 'fire', for example. Back in St Andrews, between 1738 and 1739 (the last few years of his short life), Martine carried out experiments on rates of heating and cooling and found that:

> contrary to all our fine theory, Quicksilver, the most dense ordinary fluid in the world, excepting only melted Gold, is, however, the most ticklish next to Air; it heats and cools sooner than Water, Oil, or even rectified Spirit of Wine itself.[12]

While Boerhaave's controversial distribution theory was not adopted by many investigators, Boerhaave's overall conception of a subtle heat-fluid *was* taken up and was the dominant view by the middle of the eighteenth century. These subtle fluids were proliferating—there was one for combustion (phlogiston, as above), and others for electricity, magnetism, and gravity. All were subtle (invisible) and some were imponderable (weightless) and elastic (self-repelling). Eventually, the subtlety, levity, and imponderability of these elastic fluids became too much for the scientific community to ponder. However, 'eventually' was a long time in coming (we shall return to the cases of caloric and phlogiston later) and, in the case of electricity, it could be argued that the subtle-fluid theory never went away.

The Kinetic Theory

We now turn our attention back to the mechanical or motion theory of heat. The first attempt to describe mathematically the connection between heat and the motion of the small parts was made by Jakob Hermann (1678–1733) from Basle in Switzerland, in his book *Phoromania*,[13] published in 1716. (The word 'phoromania' comes from the Greek root *phoros*, meaning 'motion'. It seems to have been a fashion in the eighteenth century to use words from antiquity.)

Hermann considered the case of elastic fluids and put forward the following prescient proposition:

> Other things being equal, heat is proportional both to the density of a hot body and to the square of the agitation of its particles.[14]

'Other things being equal' meant that this applied to all 'bodies of a similar texture'. Hermann also went on to explain that 'The agitation is the *mean* of the individual speeds'[15] (emphasis added). This remarkable insight was not taken up again until Maxwell some 150 years later (see Chapter 17).

The next investigation into the motion theory of heat as applied to a gas was by Daniel Bernoulli, son of Johann. Daniel published his great work, *Hydrodynamica*,[16] in 1738 in Strasbourg, although the work itself was carried out while Daniel was at the Russian Academy of Sciences at St Petersburg, at the invitation of Peter the Great. Hermann also spent some years in St Petersburg, and Daniel and Jakob Hermann would certainly have had many useful discussions together. This was in contrast to the help that Daniel could expect from his father. Johann Bernoulli not only went out of his way to block his son's chances of getting a professorship in mathematics at Basle, but published his own work on mathematical physics, *Hydraulica*, and predated it so that it looked as if the work was completed before Daniel's masterpiece.

Despite Johann's efforts, *Hydrodynamica* sealed Daniel Bernoulli's reputation across the whole of Europe. (It includes the effect for which Daniel Bernoulli is famous today—the different rates of fluid flow for different routes and which can result in 'lift'.)

In Chapter X of *Hydrodynamica*, Daniel gives his kinetic theory of gases. He considers the 'elastic fluid' (air) to be contained within a cylinder with a weighted piston at the top (Fig. 5.2). The fluid contains 'very minute corpuscles, which are driven hither and thither with a very rapid motion'. (The expression 'hither and thither'[17] is the only hint that the speeds and directions are randomly distributed.) The pressure of the air is due to the impacts made by these corpuscles with the walls and piston of the cylinder (this explanation of pressure is generic to all subsequent kinetic theories of gases). When the volume of the air is reduced by pushing down on the piston, then 'it appears that a greater effort [pressure] is made by the fluid for two reasons: first, because the number of particles is now greater in the ratio of the space in which they are contained and, secondly, because each particle repeats its impacts more often'.

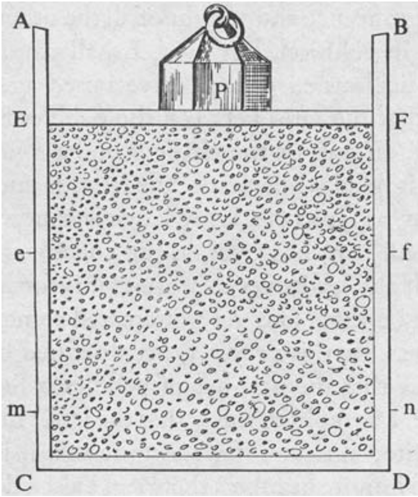

Fig. 5.2 Daniel Bernoulli's kinetic theory of an elastic fluid, from *Hydrodynamica*, 1738 (by permission of Dover Publications, Inc.).

According to Bernoulli, if the volume, V, is reduced to V/s, then the number of particles 'contiguous' to the piston will be increased by $s^{2/3}$ and the mean separation of the particles will be decreased by $s^{1/3}$. This will result in a pressure increase from P to Ps—and thus the product, pressure times volume, will remain unaltered. This is Boyle's or Mariotte's Law, a law that is confirmed by 'manifold experience'. Bernoulli is admirably cautious and adds 'whether it [Boyle's Law] holds for considerably denser air I have not sufficiently investigated' and notes that 'the temperature of the air while it is being compressed must be carefully kept constant'.

He goes on to note that heating the air will increase the internal speeds, v, of the particles and will lead to an increase in the pressure, P, on the piston: 'it is not difficult to see that the weight P should be in the duplicate ratio of this velocity [$P \propto v^2$] because, when the velocity increases, not only the number of impacts but also the intensity of each of them increases equally'. In other words, Amontons' experimental finding, that pressure is proportional to temperature, is given a theoretical underpinning.

Daniel Bernoulli's novel and essentially correct description of a gas was almost totally ignored for over 100 years. (I say 'almost' because his work was picked up by the Swiss school of natural philosophy and was an important influence on Pierre Prévost—see Chapter 6.) Nowadays, Bernoulli's description of a gas as particles in rapid, random motion feels completely intuitive, especially in comparison with a 'passion-fruit jelly'—but intuition changes with the centuries.

The modern kinetic theory incorporates some aspects that were totally foreign to the eighteenth-century mind.

First, Bernoulli's analysis did not consider a spread of particle speeds (at a given temperature). Even though Bernoulli was one of the founders of probability theory, and he advised Catherine the Great on smallpox innoculation, the idea of a *distribution* of speeds was too advanced. (One can, nevertheless, glimpse the concept of *random* motion when Bernoulli describes the corpuscles as moving 'hither and thither'.)

Also, the passage of sound was actually easier to understand in a jelly: in the kinetic theory, one needs to understand the tricky (statistical) concept of 'mean free path' and that this path-length must be much less than the wavelength of sound. (Curiously, the erroneous material theory of heat, in the hands of Laplace, was later able to account for sound quantitatively and correctly; see Chapter 10.)

Thirdly, the kinetic theory found it harder to explain the transmission of heat from the Sun to Earth across vast tracts of empty space. (The material theory also struggled with this, and Boerhaave admits that 'it appears very extraordinary that the sun, after a continual emission of corpuscles of fire upwards of 5,000 years, should not yet be exhausted'.[18])

Finally and most importantly, a conservation principle for heat didn't follow automatically from the kinetic theory but seemed quite natural in the material theory (heat would be conserved if the 'substance of heat' was conserved). True, the heat-fluid was 'subtle' (invisible) and 'imponderable' (weightless), but then the particles of a gas were also subtle and individually weightless as far as the technological capabilities of this era were concerned.

That the wrong theory should prevail is not so unusual, but what is remarkable is that there were so few tell-tale clashes with phenomenology: after all, a gas does *not* approximate to a 'passion-fruit jelly'. While pausing (until Chapter 8–12) to consider what these clashes may have been, it is interesting to observe that there have been other famous theories that were 'wrong' and yet were not in great conflict with experiment. Perhaps the most famous was the Ptolemaic system for the motion of the planets. These were said to travel around a fixed Earth in circular orbits modified by epicycles, deferents, and so on: the actual motions were described correctly. Sticking with the celestial scenario, Einstein's General Theory of Relativity took a radically different approach from Newton's theory of Universal Gravitation, yet the experimental differences between these two theories were very slight (for example, both agree on the Moon's orbit around the Earth).

One thing that Daniel Bernoulli did achieve, albeit without due recognition, was the introduction of the conceptual tool that would dominate heat and gas studies for over 100 years, right up to the formulation of the laws of thermodynamics in the middle of the nineteenth century. This was the scenario of a gas or vapour trapped in a cylinder and enclosed by a piston. It was no accident that Bernoulli had incorporated this scenario into his kinetic theory; it was a feature of the new 'fire-engine' technology (as in Papin's suggested engine at the end of the seventeenth century; see Chapter 4) and Bernoulli was certainly very interested in this. He had heard rumours of a revolutionary new engine in England, most probably Newcomen's engine. This engine had everything to do with heat and gas, and it is with this new invention that we end our coverage of heat and gas studies during the first half of the eighteenth century.

The Newcomen Engine

Not much is known about the life of Thomas Newcomen (1663–1729), inventor of the first true steam engine. He was an ironmonger and blacksmith from Dartmouth in Devon, a member of the Calvinist sect of Anabaptists, and he died in London, probably of the plague.

Newcomen's first acquaintance with the steam engine came when he was called upon to advise on problems with a Savery engine that was being used to pump water out of the mines in neighbouring Cornwall. Newcomen and his assistant, the glazier John Cawley (or 'Calley'?), carried out a series of tests on Captain Savery's engine. These tests led Newcomen to make some improvements to the engine which, taken individually, were impressive enough but, taken together, were momentous and resulted in the invention of the first steam engine that actually worked.

The most noticeable difference between Newcomen's 'fire-engine' and Savery's 'water-commanding engine' was that the steam chamber in Newcomen's engine was now a cylinder with a piston (Fig. 5.2). The piston was connected by a chain to one end of a great beam that oscillated about a central pivot. The other end of the beam was connected to a counterweight—the mine pumping gear itself. A boiler, directly under the cylinder, filled the cylinder with steam. At the same time, the descending counterweight pulled the piston up and allowed the steam to expand. (The steam pressure was not very high and barely contributed to driving up the piston.)

A second difference between the Newcomen and Savery engines was the introduction of various valve mechanisms, which led to a more automated engine. Thus, when the piston reached the top of the cylinder, the great beam caused two valves to trip, one to cut off the supply of steam and the other to switch on an internal spray of cold water (the latter causing the steam to condense). The steam condensed and the warm condensate and condensing water were collected through an 'eduction' pipe, another novelty, and returned to the boiler. As the steam condensed, a partial vacuum was created under the piston. Atmospheric pressure then drove the piston down and the counterweight was brought back up. This constituted the power-stroke of the engine. When the piston reached the bottom of the cylinder, the two valves were tripped again, one switching off the spray of cold water and the other turning on the supply of steam—and the cycle recommenced.

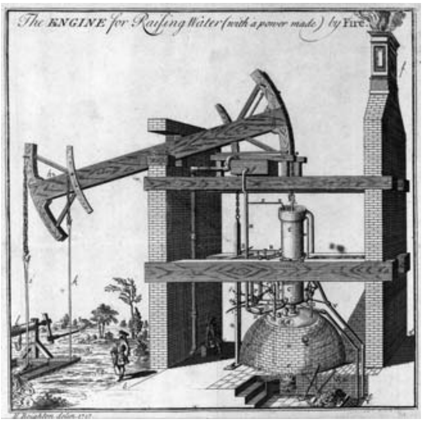

Fig. 5.3 Newcomen's 'fire-engine', Oxclose, Tyne & Wear, 1717 (Science and Society Picture Library, Science Museum, London).

Timing was everything: the great lumbering beam could move only slowly, yet the valves had to be tripped with snap-action precision. As Professor Cardwell says, Newcomen's valve mechanism represented 'a display of inventive ability amounting to genius'.[19] Prior to Newcomen, all engines required the continual presence of operators to start and stop the supplies of steam, cold water, and so on. For the first time a steam engine could now be truly self-acting.

The final modification introduced by Newcomen combatted the problem of dissolved air. This ended up in the cylinder and no amount of cold water would make it condense (the Savery engine regularly stalled on this account). Newcomen invented the 'snifting valve', which allowed the engine to 'sniff' away the unwanted air.

Newcomen knew nothing of the various theories of heat and gas, but through his superb engineering intuition he invented an engine that was of as much historic importance as the Gutenberg press and the weight-driven clock.[20] The influence of the engine was unstoppable and it spread rapidly to all the mining areas of England (it was first used in Dudley in 1712) and then on to the continental mines of Dannemora (Sweden), Schemnitz (now Slovakia), Passy in France, and on to North America. While the physics of the engine was not well understood (for example, it was still not known that all matter exists in one of three states—solid, liquid, or gas), it was appreciated for the first time that *heat* ('fire') was the prime mover of the engine. All previous working engines had employed water, wind, or muscle. (Savery's engine barely worked and Amontons' and Huygens' engines were just proposals.) The success of the 'fire-engine' could not be questioned; the physics would just have to catch up.

6

The Discovery of Latent and Specific Heats

Throughout the whole of the eighteenth century Paris was the undisputed centre of science, even during the French Revolution (in 1789). There were other lesser centres, such as Basle and Geneva in Switzerland, St Petersburg in Russia, and Stockholm in Sweden. After Boerhaave, Dutch science waned and German science was at its lowest ebb.

In Britain, Newton was a hard act to follow. The Royal Society in London degenerated into something akin to a gentleman's club and the universities of Oxford and Cambridge were moribund as far as the study of natural philosophy was concerned. The action moved north to Scotland (Scotland had five universities to England's two) and then on to Birmingham and Manchester.

Much has been written about the rise of the Scottish school of science and about what became known as the Scottish Enlightenment. However, in the case of the discovery of latent and specific heats, climate was probably as influential as any other factor; Joseph Black and Johan Wilcke came from the cold of Scotland and Sweden, respectively, and both men made their discoveries after everyday observations of snow and ice.

Joseph Black

Born of Scottish parents in the French province of Bordeaux (his father was a wine merchant) and educated at a private school in Belfast, Joseph Black (1728–99) advanced our understanding of heat more than any other philosopher in the eighteenth century. Aged only 16, he went to study medicine at the University of Glasgow in Scotland.

Black soon found himself more interested in chemistry than in medicine. The turning point came when he started to attend the chemistry lectures of William Cullen (1710–90). Cullen was a doctor and was founder of the science of nosology. He was quite a character 'known everywhere for his huge peruke, bigger hat, big coat-flaps sticking out and huge sand-glass to measure patient's pulses'.[1] Quite quickly, the relationship between Cullen and Black developed from master and pupil to professor and valued assistant, and then to a life-long friendship. In their Edinburgh days they were to meet at the Oyster Club, a weekly dining club started off by Black and two other friends, the economist Adam Smith and the geologist James Hutton—all figures in the Scottish Enlightenment. They also met at the Poker Club where, over a sherry, a shilling-dinner, and claret, members were 'roasted' for their views.[2]

Black's dissertation for his MD at the University of Glasgow was nominally an examination of the treatment for urinary calculi ('stones'), but led Black to a major discovery—he was the first person to reform the generic mixture 'air' into chemically distinct components by identifying one of these—'fixed air' (what we now call carbon dioxide). He did this by perfecting the chemical balance and using it more systematically than any chemist had done before.[3] He was thus able to attribute the decrease in weight of *magnesia alba* (magnesium carbonate) to the expulsion of 'fixed air' rather than to the gain of a causticizing agent, the negatively weighted phlogiston.

Black's other major discoveries were in the field of heat, but all shared this feature of respectful attention to quantitative measurements. This was more inclined to lead to knowledge and understanding than 'vague reasoning'.[4] The influence of the philosopher David Hume is evident. (Hume was a leading figure in the Scottish Enlightenment—also a member of the Poker Club and a friend and patient of Black's.)

General Properties of Heat

Black recognized the crucial distinction between heat and temperature:

> Heat may be considered, either in respect of its quantity, or of its intensity. Thus two lbs. of water, equally heated, must contain double the quantity [of heat] that one of them does, though the thermometer applied to them separately, or together, stands at precisely the same point.[5]

This still didn't reveal how much heat, in an absolute sense, there was in just one pound—or any other amount—of water. (If one could count the individual particles making up the heat-fluid—if there *was* a heat-fluid—then presumably one would have an absolute measure of heat.) Black, however, was content to adopt a purely operational procedure, and in his later measurements the heats were always relative to temperature changes in a given mass of water.

Black read more significance than anyone else into the common observation that different bodies, left to themselves, will all acquire the same temperature:

> if we take one thousand, or more, different kinds of matter, such as metals, stones, salts, woods, corks, feathers, wool, water and a variety of other fluids, although they be all at first of different heats, let them be placed together in the same room without a fire, and into which the sun does not shine, the heat will be communicated from the hotter of these bodies to the colder, during some hours perhaps,…, at the end of which time, if we apply a thermometer to them all in succession, it will point precisely to the same degree.[6]

He elevated this observation to the status of a law:

> We must adopt, as one of the most general laws of heat, the principle that all bodies communicating freely with each other, and exposed to no inequality of external action, acquire the same temperature as indicated by a thermometer.

He realized that this was really quite a remarkable fact, not suggested by any properties of the individual bodies:

> No previous acquaintance with the peculiar relation of each body to heat could have assured us of this, and we owe the discovery entirely to the thermometer.

He further understood that this equality of temperature was not the same thing as an equality of heat, but represented an '*equilibrium* of heat' between the various bodies.

This idea of bodies arriving at a common temperature is something we are very familiar with—so familiar, in fact, that it is hard to see that it is not an obvious result. However, bodies are generally *not* found to be all at the same temperature: people and animals maintain their own specific temperatures, marble may feel cold, metals may feel hot or cold to the touch, the ground is warmer than the air temperature at night, lakes and ponds are often warmer or colder than their surroundings, and so on. To recognize that all bodies are striving to reach the *same* temperature was a major achievement. However, Black did miss one important and 'obvious' result—that bodies are generally striving to cool down rather than heat up.

Latent Heat

As we have seen, Black made the crucial distinction between temperature and quantity of heat. He also carried out a series of ingenious and simple mixing experiments in order to be assured that each degree of the thermometer represented an equal quantity of heat. (This method of mixtures had already been used by Brook Taylor in 1723; see Chapter 5.) It was this understanding of the difference between heat and temperature that led to Black's first discovery in the science of heat, the discovery of the concept of latent heat. Once again, Cullen was the starting point.

Cullen had noticed a surprising and inexplicable phenomenon—the intense cold produced when 'spirituous' liquids, such as ether, were allowed to evaporate, especially when this occurred in an evacuated receiver. He communicated these findings to Black. At the same time, Black was perusing Boerhaave's *Elementiae chemicae* (*Elements of Chemistry*; see Chapter 5) and was intrigued by a description of one of Daniel Fahrenheit's experiments. Fahrenheit had managed to cool a sample of water some 8 degrees* below the freezing point (32°F). The water remained liquid (we would now say that it was supercooled) provided that the vessel was small and was left completely undisturbed. The slightest agitation, even air currents on the surface of the water, resulted in the formation of feathery ice crystals spreading rapidly in a network. Most intriguing of all, as the water froze its temperature jumped discontinuously from 24 °F to 32 °F. As Black put it, the 'heat unnecessary to ice' was liberated from the water.

Black mused on these findings and also on his observations of the snow around Glasgow. When the weather grew milder and the temperature rose above the freezing point, Black noted that this did not result in an immediate thaw of all the snow:

> were the ice and snow to melt... suddenly, the torrents and inundations would be incomparably more irresistible and dreadful. They would tear up and sweep away everything, and that so suddenly, that mankind should have great difficulty to escape from their ravages. This sudden liquefaction does not actually happen; the masses of ice or snow melt with a very slow progress... [which] enables us to preserve it [ice] easily during the summer, in the structures called Ice-houses.

* The temperatures in this chapter are all on the Fahrenheit scale.

Black saw that the very slow melting of snow, the cooling when nitrous ether evaporates, and the heat liberated in Fahrenheit's experiments all pointed to the fact that substantial amounts of heat were involved. These quantities of heat were specifically for the purpose of 'rendering the ice fluid', or converting liquid into vapour—in other words, for changes of state. What's more, the heat that enabled a change of state did not cause any change of temperature. Amontons had already shown that boiling water kept a constant temperature; but while the temperature remained constant, a transfer of heat *was* taking place, as evidenced by the continuous supply of heat being drawn from the boiler. Black called the heat required to enable a change of state, but not showing up as a change in temperature, the hidden or 'latent' heat.

Black's introduction of the concept of latent heat presupposed a commitment to the *conservation* of heat. Rather than saying that heat simply disappears during a change of state, he held that heat *is* conserved but becomes latent.

It was not sufficient to argue just from 'vague reasoning', experimental confirmation was required. Black carried out the experiments in the large hall adjoining his college rooms, as it was isolated and sufficiently warm (47°F was evidently considered warm in eighteenth-century Scotland[7]). A sample of ice and an equal mass of ice-cold water were observed as the former thawed and the latter heated up:

> I then began to observe the ascent of this thermometer, at proper intervals, in order to learn with what celerity the water received heat, stirring the water gently with the end of a feather about a minute before each observation.[8]

The time taken for the ice to just melt away completely was measured (it took 10½ hours) and this was compared to the time taken for the water to warm up from 33°F to 40°F (half an hour). Assuming that the room's temperature and other factors had remained constant for the 10½ hours, Black took it that the ice had received 21 times as much heat as the water had received.

Black also attempted the reverse experiment, namely, a measurement of the heat given up when a certain amount of water freezes. However, it proved much harder to find a source of cold that was reliably constant—there was no inkling at this stage in history that there should be any asymmetry between the processes of cooling and heating.

Black soon extended his ideas to the case of the vaporization of water. As before, he argued that the change of state couldn't happen instantaneously, or 'water would explode with the force of gunpowder'.[9] He attempted to obtain supercooled steam but it was too tricky. He then fell back on the usual methods: he measured the time required to just boil away a given mass of water and the temperature increase in another equal mass of water during the same time interval. He asked a 'practical distiller'[10] about the constancy of his furnace. The distiller assured Black that he could tell, to a pint, the quantity of liquor he would get in an hour. Black checked up on this with his own laboratory furnace, seeing how long it took to boil off small amounts of water.

Black found that the heat absorbed to convert water into vapour was equal to that which would have raised the temperature of an equal mass of water by 810°F—if the water hadn't all boiled away first (this is about 17% lower than the modern figure). He stopped these experiments for a couple of years but then, in 1764, came a great stimulus to further enquiry; this was the year in which James Watt discovered a new, improved version of the Newcomen steam engine. We shall take up the threads again in the section on Watt in Chapter 8.

Specific Heat

As with Black's discovery of latent heat, his discovery of specific heat (he calls it 'heat capacity') was started off by his reading about an experiment of Fahrenheit's, again in Boerhaave's *Elements of Chemistry*. Fahrenheit had mixed water and mercury together and found that in order to obtain a final temperature equal to the mean, he needed to mix three volumes of mercury to every two volumes of water (Chapter 5). We saw how this had led Boerhaave to conclude that heat is distributed 'in proportion to [a body's] extension'. This was something that Black could not agree with, especially as Fahrenheit's experiment appeared to contradict it (there was a 50% greater volume of mercury than of water).

Black was trying to mesh Fahrenheit's findings with those of George Martine in 1740. Martine had found that:

> spirit of wine both heats and cools faster than water, and that in a much greater proportion than the inverse ratio of their Specific weight does require, as we observed likewise of oil. But still quicksilver, however dense, is more ticklish and easier affected by heat and cold than any of these fluids. Common brandy, upon trial, I did not find to differ sensibly from water in this respect.[11]

It was suddenly clear to Black how to explain both Fahrenheit's and Martine's results—by supposing that mercury has a smaller store of heat than an equal mass of water. A greater volume of mercury would then be required in Fahrenheit's mixing experiment and, in Martine's experiment, a smaller quantity of heat would be implicated for mercury, and therefore a given temperature change would occur more quickly for mercury than for water.

Black went on to assert that all other substances also had their own characteristic capacity for heat—but he was not able to fully justify this assertion. Take the experiment of placing a cube of iron and a cube of wood in the same oven for an equal length of time. Upon removal from the oven, the iron felt much hotter, and so, Black argued, its capacity for heat was the greater. In fact, *wood* has a greater heat capacity, but its conductivity is much lower than that of iron. Black didn't take account of factors such as conductivity, emissivity, and so on. He thought that the actual speed of heat entering a body was the same for all materials—except, apparently, for those of a spongy consistency.

Apart from a few desultory measurements, Black never carried out a systematic survey of specific heats and he never wrote up any of his ideas. It was left to his students to publish his lecture notes and also to defend his priority. By all accounts Black was a brilliant lecturer, mostly *ex tempore*, and with many experimental demonstrations to help develop the argument. A student from Black's Edinburgh years described him as follows:

> He wore black speckless clothes, silk stockings, silver buckles, and either a slim green silk umbrella, or a genteel brown cane... The wildest boy respected Black. No lad could be irreverent towards a man so pale, so gentle, so elegant, and so illustrious. So he glided, like a spirit, through our rather mischievous sportiveness, unharmed.[12]

Johan Wilcke

While studying the shapes of snowflakes and ice crystals in Sweden in 1769, Johan Wilcke (1732–96) made the observation that water that was cooled below 32 °F became warm on freezing. In the winter of 1772 (so the story goes), Wilcke was trying to sweep up snow from a small courtyard and, tiring of the broom, he tried melting the snow with hot water. He was surprised at how ineffective the hot water was—a large amount only melted a small quantity of snow.[13]

In this way, around ten years after Black and quite independently of him, Wilke came to the same conclusions from almost the identical evidence, and formulated the concepts of latent and specific heat all over again. As, unlike Black, Wilcke published his work, he got much of the credit.

(The Finnish scientist, Johan Gadolin (1760–1852), continued Wilcke's work on specific heats. The element gadolinium was named after him.)

Irvinism

Black's work on specific heats was extended by one of his students and a future professor at Edinburgh, William Irvine (1743–87). Irvine had the following ingenious idea: all bodies contain a certain absolute quantity of heat, dependent only on their temperature and their capacity for heat. Just as a tank may hold a certain volume of water dependent on its cross-section and its height, so a body may hold a certain volume of heat 'poured in' up to a certain level. The fluid theory of heat is virtually written into Irvine's model.

Irvine argued that any change in specific heat capacity would lead to a change in temperature and vice versa. In other words, Irvine didn't believe in latent heat. He had it that heat used up or given out during a change of state or chemical reaction was due solely to changes in the heat capacities of the reactants or the products. As specific heats could be measured, this laid Irvine's theory open to experimental test.

Irvine's theory threw up another interesting possibility—what if the tank was empty? This would represent a state of zero heat and would occur at the absolute zero of temperature. Predictions of this absolute zero provided another quantitative check of Irvine's theory.

Antoine-Laurent Lavoisier and Pierre-Simon Laplace, in their work on heat in 1783 (see Chapter 8), found that Irvine's theory wasn't borne out either by the data on specific heat capacities or by the predictions for absolute zero, the latter varying wildly from −600 °C to−14,000 °C.

Like Black, Irvine did not like to put pen to paper, and it was left to his son to tidy up his father's papers and publish his work posthumously.

The Specific Heats of Gases

The last of the Scottish school to continue Black's work on specific heat was Adair Crawford (1748–95). He was a surgeon at St Thomas' Hospital in London, but he visited Scotland to learn about the latest theories on heat and attend some of Irvine's lectures. He had a goal: he wanted to explain the perplexing problem of animal heat. In Glasgow, in the summer of 1777, he measured the heat capacities of a variety of 'animal, vegetable and mineral substances'[14] and—a first—of 'atmospherical and fixed air'. He used a thin hog's bladder, washed it with soap and water, dried it, and forced any residual air out by rolling packing thread over it. After the given 'air' had been introduced into the bladder, it was heated in a Dutch oven, wrapped in flannel to keep it warm, and then pressed out of the bladder using weights. The gas was forced through a worm (a crooked glass tube) containing half an ounce of cold iron filings and the temperature rise of these was measured.

Needless to say, the experiments were not very accurate. They were repeated again later using a brass cylinder to contain the gas and a water-bath for the heat to be transferred to. However, the results in both sets of experiments showed what Crawford wanted them to show—the heat capacity of the 'fixed air' (the 'air' that is breathed out) was considerably less than the heat capacity of 'atmospherical air' (the 'air' that is breathed in). This, Crawford presumed, was the source of animal heat.

Unwittingly, Crawford had maintained the gas at a roughly constant pressure in the bladder and at a constant volume in the brass cylinder. This distinction was to be important in the future study of the specific heats of gases (see Chapter 10), but Crawford was unaware of it.

Radiant Heat

An obscure apothecary in Sweden discovered that there were two kinds of heat, never mind two kinds of heat theory. This was Carl Wilhelm Scheele (1742–86) from Stockholm, Uppsala, and, finally, Koping (on his deathbed in Koping, Scheele quickly married the pharmacist's daughter so that she would inherit the pharmacy).

(The German autodidact and polymath Johann Heinrich Lambert (1728–77) was another to discover 'obscure heat',[15] as he called it. In the eighteenth century, it was still possible for a very gifted and hard-working individual to encompass the whole sweep of human knowledge. Lambert was such a one, and he certainly was hard-working: he worked from 5 a.m. to midnight with only a two-hour break at noon.[16] He is famous for his photometer, for his proof that π is irrational, and many other things.)

Scheele was an outstanding chemist who discovered oxygen (before Joseph Priestley), chlorine (before Humphry Davy), and many other chemicals besides. Recently (1982) there has been speculation that the pigment known as Scheele's Green (copper arsenite) may have been implicated in Napoleon's death (the wallpaper in Napoleon's room on Elba was green).

By careful observations and clear thinking about everyday phenomena, Scheele identified 'radiant heat' and distinguished it from 'familiar heat' and from light.[17] The observations required no special equipment and could have been made centuries earlier—but they weren't.

Scheele noted that warmed air rose upwards from a hot body, such as an oven, and that you could see the air shimmering where convection currents (modern expression) were occurring. However, a completely different sort of heat travelled in rays in all directions (not just upwards) and didn't involve the air. In fact, Scheele found that the air could be disturbed, as in a draught, and the 'heat-rays' were unaffected. Also, a candle burning and smoke ascending were not deflected when crossed by such rays; and that while the shadow of a window frame didn't quiver, the shadow of a hot iron or stone did quiver.

This new heat (Scheele coined the term *radiant* heat) was similar to light: it travelled in straight lines and followed the optical laws of reflection. Indeed, heat sources were usually also light sources, as in the prime example of the Sun. But Scheele saw that radiant heat was not only different from convective heat, it was different from light as well. Light was reflected from a polished metal surface and was transmitted through and reflected by glass, while radiant heat was reflected by the metal but *absorbed* by the glass—'remarkable', commented Scheele. Scheele also noted that the greatest radiant heat was obtained not from a brightly burning fire, but from the glowing charcoal left behind. Also, a fire could be seen from far away, but its warmth felt only from nearby ('at about three ells' distance'). (An ell is the distance from shoulder to fingertips.)

A hot body was just 'hot', whatever kind of heat was used to warm it. Also, the different kinds of heat could be converted into and separated from each other. For example:

> at two ells' distance in front of the stove, by means of a small metallic concave mirror, a focal point can be produced which kindles sulphur. Such a mirror can be held for a very long time in this position without its becoming warm; but if it is coated with some soot over a burning candle, it cannot be held for four minutes in the former position before the stove without burning the fingers upon it...

and

> the metallic concave mirror and the plate of metal rapidly become hot [by conduction] as soon as they touch a hot body, although they do not become the least warm from the [radiant] heat proceeding from the stove.

Scheele's work on radiant heat was continued by the Swiss school at Geneva, especially Horace-Bénédict de Saussure (1740–99), Marc-Auguste Pictet (1752–1825), and Pierre Prévost (1751–1839). These scientists were part of a new breed—mountaineers. They explored the rugged Alpine landscapes for fun (a new Romantic idea and a far cry from Thomas de Quincy, who lowered the blinds of his coach to shut out the Lake District views[18]). De Saussure was one of the first to climb Mont Blanc. The mountains were nature's laboratory and led to advances in geology, meteorology, and associated physics.

Pictet was the first philosopher to attempt to measure the speed of radiant heat, but could only conclude that it travelled 'perhaps as rapidly as sound or even light'.[19]

De Saussure and Pictet showed that a hot, but not glowing, bullet at the focus of one concave metal mirror produced a rise in the thermometer placed at the focus of another mirror. In order to prove that heat rather than light was involved, Pictet repeated the experiment with a non-luminous source—a glass vessel of boiling water. Once again, the 'obscure heat' was detected and brought to a focus at a secondary mirror.

Pictet went on to repeat the arrangement yet again, but this time placing *ice* at the focus. Even so, Pictet was sceptical, saying that 'cold was only privation of heat' and that 'a negative could not be reflected'.[20] He found, however, a positive result—the temperature at the second mirror *was* lowered. (This was the same effect as had been examined by the Accademia del Cimento in the seventeenth century; see Chapter 4. It can be demonstrated fairly easily at home—see the advice and explanation given in Evans and Popp.[21])

Pictet's apparent radiation of cold caused a stir amongst natural philosophers. Its chief benefit was to provoke the physicist and literary figure Pierre Prévost (who had translated Euripides) into examining it.

According to Prévost, the heat-fluid (by now, 1791, called 'caloric') was made up of discrete heat-particles moving rapidly in all directions. Thus, while Prévost assumed the mainstream caloric theory, it's clear that he was also drawing from the motion theory and from the work of his countryman, Daniel Bernoulli, published half a century earlier.

The crucial aspects of Prévost's theory were as follows: the heat-particles are tiny, almost infinite in number, and move in all directions; a body at a higher temperature has faster heat-particles and therefore emits caloric at a higher rate; and all bodies are at the same time continually emitting and absorbing heat-particles to and from their surroundings.

As a hot body emits caloric at a higher rate than a cooler body, the hot body will emit more caloric than it receives and will cool down. The cooler body will emit less caloric than it receives and will therefore heat up. When the rate of emission and absorption are equal, the body will maintain a constant temperature. This was Prévost's 'Theory of Exchange'[22] or concept of 'mobile equilibrium'.

Prévost's theory explained how bodies could cool down, but what did it have to say about Pictet's experiment, about the transmission of cold? In fact, Prévost had shown that the transmission (and reflection) of 'cold' is impossible. A body will only cool down if it is hotter than its surroundings; in other words, a body can never be the net recipient of 'cold'. This asymmetry between 'hot' and 'cold' passed without a murmur, but was to have big repercussions 50 years later: it would lead to a new law of thermodynamics.

Pictet remarked that Prévost's theory might be applicable to the 'electrostatic machine', as this simultaneously acquired and discharged electricity. Another link between heat and electricity was made by the Dutch scientist Jan Ingen-Housz (1730–99). He found that good conductors of electricity were also good conductors of heat. He measured specific thermal conductivities by making up wires in various materials and coating them with wax. He then heated one end and measured the speed with which the wax 'melting point' travelled along the wire. One other relevant bit of phenomenology was that heat and electricity, and also light, were linked in another way—they could all be generated by friction.

Overview

Black established a clear distinction between heat and temperature, defined 'quantity of heat', and showed that bodies are constantly striving to achieve a common *temperature* rather than a common heat.

As the end of the century approached, the material theory of heat was in the ascendant. Lavoisier had made it one of his chemical 'elements' and called it '*calorique*' (see Chapter 8). Black professed to be undecided on the question, but privately he held to the material rather than the motion theory of heat. (Although he never admitted any connection, it is hard to imagine that he didn't draw some parallels between the containing of his 'fixed air' within a solid and the 'fixing' of the highly elastic heat-fluid in the form of latent heat.)

Radiant heat was discovered and there were more and more parallels being noticed between ordinary heat, radiant heat, light, and electricity.

Prévost's theory of 'mobile equilibrium' was a fundamental advance. It supplied a mechanism by which Black's common temperature could be achieved, but much more than this, it introduced a totally new idea—the idea of a *dynamic* equilibrium.

As is so often the case, a new idea (e.g. that of dynamic equilibrium) seems intuitive to us now, but was strange and unfamiliar in the beginning. The heat-fluid in Irvine's 'tank' finds its level, but this is a once-and-for-all process. Prévost's equilibrium was quite different, it was maintained by continual small adjustments. Count Rumford—we shall meet him in Chapter 9—thought it was ludicrous: how could a body be both emitting and absorbing something at the same time? Maxwell, on the other hand, saw Prévost's theory as underpinning his own kinetic theory of gases (see Chapters 17 and 18). Finally, Prévost's Theory of Exchange, having its roots in both the material and motion theories of heat, shows us that these theories were not as antithetical as is sometimes assumed.

The eighteenth century had a preoccupation with the 'heat-of-a-body' and this was taken to its logical extreme in 'Irvinism'. We now don't consider the 'heat-of-a-body' to be a particularly telling parameter: it isn't of universal importance for the very reason that it is too body-specific. In the next chapter, we shall find that the 'force-of-a-body-in-motion' was also a misleading concept.

7

A Hundred and One Years of Mechanics: Newton to Lagrange

Part I
The 'Force of a Body in Motion'

We will put the study of heat to one side and consider some other 'blocks' of energy that started to be recognized in the eighteenth century—the 'blocks' of mechanical energy. Mechanical energy emerged from mechanics, the study of bodies in motion. The bodies could be as small as the particles or atoms from which, some said, everything is made, or as large as a celestial body, or any size in between.

The cause of prime motion was uncertain, and generally taken to be God-given. After this, the cause of different motions was due only to the mutual interactions of the bodies. (While Newton had introduced forces, he still thought of these as ultimately due to a *body* and not an 'abstract mathematical point'.) Although it was, of course, known that a body's motion could be altered by the intervention of mind (as when I decide to pick up an apple), the mechanics was not so ambitious as to attempt an explanation of this sort of thing. In fact, such effects were rigorously excluded. According to Leibniz, even God didn't interfere after He had set everything up.

Although there were important precursors (as outlined in Chapter 3), mechanics as we know it today was essentially started off by Newton. We shall remind ourselves of Newton's three Laws of Motion:[1]

LAW I: Every body perseveres in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed thereon.

LAW II: The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed.

LAW III: To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

What Newton had done in stating his laws of motion was amazing: from the plethora of phenomena—a horse pulling a cart, a fly walking on water, magnets attracting each other, a drop of water being drawn up a narrow glass tube, the motions of the planets and of the Moon, the spinning of the Earth, the tides, a boy playing with pebbles, and so forth—to wrest out the concepts of 'nothing is happening', 'something is happening', and force. Force explains all the cases where 'something is happening' in just one way—the motion is changed from uniform to accelerated. This was all the more amazing when we remember that, apart from free fall, there were no smooth accelerations around in those days—no cars with accelerator pedals, only carriages jolting along rutted roads. Newton wasn't sure but he suspected that *all* the phenomena in nature, including chemistry and the action of light, would be explainable in this way.

You might think that all the natural philosophers would say, 'Well, that's that then. Let's all go home'—but not a bit of it. Newton *was* lauded for his Law of Universal Gravitation (except by the followers of Descartes, who deplored the implied 'action-at-a-distance'), for explaining the motion of the Moon, the figure (shape) of the planets, and the path of Halley's comet. But his Laws of Motion went almost unnoticed. Even in Cambridge, Newton's old university, the mechanics textbook used was that of the Cartesian, Jacques Rohault, albeit with Newtonian annotations.

There was, of course, the time required for the dissemination of the work. Newton's *Principia* was published in 1687, but it was only after 1727 that his work was first promoted in France, by Voltaire (1694–1778). Voltaire had judiciously absented himself from the environs of the French court after two stays in the Bastille. He visited London, where he was greatly impressed by Newton and Locke and all things English. Upon returning to France, Voltaire won over his mistress, the Marquise du Châtelet (1706–49), to Newton. She helped Voltaire with his *Elements de la Philosophie de Newton* and subsequently she translated the *Principia* into French—until recently, this was the only French version. The Italian philosopher Francesco Algarotti (1712–64) wrote *Newtonianesimo per le dame, ovvero Dialoghi sopra la luce e i colori* (*Newtonianism for Ladies, Dialogue on Light and Colour*), but this concentrated on optics. It was in

Holland where Newton was most admired and where Willem 's Gravesande published his textbook *Mathematical Elements of Physics*, which was an exposition of Newton's physics.

The *Principia* is probably the most influential book ever written (by a single author) even if it is one of the least read. However, Newton was secretive, defensive, and combative. Not surprisingly, there followed an era of conflict and debate as the world struggled to absorb the new world-views.

The chief opponent was Leibniz who had to defend himself against Newton's charges of plagiarism regarding the discovery of the calculus. As well as the calculus, the conflict was about fundamental differences in the metaphysical position of the Newtonians and the Leibnizians. It included such questions as: Were space and time absolute and universal and was space a void? Were there atoms in this void, and were they hard and impenetrable? Could forces act (and act instantaneously) across this void? Did Newton's clockwork universe need God to wind it up from time to time (to use Leibniz's metaphor)?

All these issues were interrelated but we shall concentrate on just one—what is the best measure of 'force'? This is because this debate evolved into two distinct outlooks: the Leibnizian 'energy view' and the Newtonian 'force view'.

This controversy was initially between Leibniz and Descartes (though Descartes had in fact died 36 years earlier). It was sparked by a paper published by Leibniz in 1686, the year before the *Principia* appeared. In this paper (we have covered it already in Chapter 3), Leibniz trumpeted the fact that Descartes' quantity of motion, mv, was *not* universally conserved, despite Descartes' claims, and, worse still, could lead to a perpetual motion. Leibniz claimed that his own measure for 'force' (*vis viva*, given by mv^2) *was* always conserved and did not lead to a perpetual motion.

Leibniz therefore started off the whole debate and established the two competing measures of 'force': the Cartesian mv versus the Leibnizian mv^2 (where m is the mass and v is the speed of the given body).

The Newtonians, ever ready for a fight, especially against Leibniz, were only too willing to join in the controversy. When some Newtonians, Johann Bernoulli and Willem 's Gravesande, 'deserted'[2] to the *vis viva* camp, the Reverend Dr Samuel Clarke was outraged and wrote (in 1728) that they were attempting to 'besmirch the name of the great Sir Isaac Newton'.[3] Johann Bernoulli (1667–1748) was a member of the illustrious Bernoulli family of mathematicians from Basle in Switzerland.

He was not only a 'deserter' but he became Leibniz's most ardent supporter and hostile to anything Newtonian. For example, Johann backed Leibniz in the priority dispute over the calculus and promoted Leibniz's version of the calculus on the continent.

As regards the other natural philosophers, the two camps were almost entirely split along national lines. This prompted the Scottish philosopher Thomas Reid to suggest the following absurd compromise:

I would humbly propose an amicable solution upon the following terms:

(1) In all books, writings or discourses made within Great Britain and the dominions thereto belonging…the word force shall be understood to mean a measure of motion proportional to the quantity of matter and the velocity of the body on which that force is impress; unless where the contrary is expressly declared.

(2) In all other places the word force shall be understood to mean a measure of motion proportional to the quantity of matter and the square of the velocity.

(3) All hostilities between mathematicians on both sides shall cease from the time above specified.[4]

While the controversy between Leibniz and the Cartesians was thus extended to include Newton, the Newtonians still held to the Cartesian measure of force, mv, known as momentum, albeit with the crucial modification that it now had direction as well as magnitude. This may seem puzzling to the modern reader—why wasn't the force given by $F = ma$, as defined by Newton himself in his Second Law of Motion? (Here, m is the mass and a is the acceleration.) There are a number of answers to this.

First, the quantity 'mass × velocity' had a long history, stretching back to antiquity. It expressed the condition of equilibrium in the lever of Archimedes (the masses at each end were in inverse proportion to their velocities when the lever moved). It was extended by Jordanus de Nemore in the thirteenth century, by Stevin in the sixteenth century, and by Galileo in the seventeenth century to be the condition of equilibrium for all the simple machines (inclined plane, pulley, screw, wedge, etc.) before being taken up by Descartes.

Secondly, Newton's own definition of force in his Second Law was not in fact the familiar $F = ma$. He defined force as the change in a body's momentum, but with no mention of the rate or time interval involved (Chapter 3). In fact, it wasn't until 1736 that Newton's force was properly defined (by the Swiss mathematician Leonhard Euler; see Part II below) and even then, not given in the modern vectorial form, $\mathbf{F} = m\mathbf{a}$.

Surprising as it may seem, another reason (I believe) why force was not given as **F** = m**a** is that there were few easily observable confirmations of this relationship, apart from the case of freely falling bodies.

Newton had worked out the case of the Moon falling to Earth under gravity ('the only problem which gave me a headache'[5]) but the motion was circular and its physical interpretation was hard. The attraction between floating cargoes of loadstone and iron would soon be damped by the water. There was Hooke's Law for springs and also the motion of vibrating strings—but the motion was rapid and oscillatory. None of the above cases was such as to bring out the simplicity of **F** = m**a** (I push something and it goes faster).

All this is yet further testimony to Newton's genius—that he could establish a law for which there was no pressing *experimental* need or evidence.

The concept of acceleration, let alone force, was mathematically sophisticated, abstract, and difficult. The idea that anything as cosmically significant and primitive as the 'force of a body in motion' could have a form as complicated as '*ma*' was unacceptable to some.

So let's see how mv^2 and mv shape up. The testing ground was initially collisions between bodies, as these were at the heart of the mechanical philosophy.

The findings were as follows: in the case of elastic collisions both mv^2 and mv were conserved, whereas for inelastic collisions only mv was conserved. 'Elastic' was a word that started to be used in the late seventeenth century. With reference to a fluid, 'elastic' (from the French word *elater*) meant expanding to fill its container; while with reference to a spring or a body it meant able to take up its original shape after compression. 'Elastic' as applied to a collision was taken to mean that the participating bodies were, confusingly, either absolutely hard or absolutely deformable (i.e. elastic in the sense of the spring which can regain its shape).

Jumping ahead to the modern definition of an 'elastic collision', we are plunged into tautology: the collision is elastic if the total kinetic energy is conserved; total kinetic energy is only conserved if the collision is elastic.

The definitions—both then and now—are only saved from circularity when it is realized that there are *other* ways in which elastic and inelastic collisions can be differentiated apart from the constancy of mv^2. In elastic collisions, the bodies emerge with the same shape and mass as they had before the collision, whereas for inelastic collisions the bodies may have their shapes permanently deformed, end up being joined together, or be shattered into smaller pieces.

These distinctions may seem obvious to us now, and some of the philosophers of the day (eighteenth century) thought so too—too obvious, they thought, for it to be necessary to carry out actual experiments. (Some, such as Jean le Rond d'Alembert, even thought that mechanics was a branch of mathematics and was therefore not 'contingent' on the results of experiments.) But the differences between elastic and inelastic collisions *are* contingent, and further enlightenment would only come when experimental as well as mathematical investigations were carried out.

However, checking up experimentally on collisions was no trivial matter—how were constant initial velocities to be achieved and how were final velocities to be measured? How were friction and air resistance to be sufficiently minimized and how were deformations to be gauged? There were no air tables or electronic timers. Even an accurately flat surface was beyond the casting and planing techniques of the day.

Willem 's Gravesande (1688–1742) was one natural philosopher who did go to great pains to carry out the experiments. He was famous at Leyden for his lecture demonstrations of various physical principles. (We have met him before in Chapter 2. It was he who went to visit the Landgrave of Hesse at Kassel to check up on Orffyreus' perpetual motion machine.)

's Gravesande's allegiance was initially to mv and the Newtonian school. However, after carrying out various experiments he broadened his view to include mv^2 as a useful measure (much to the disgust of the Reverend Dr Clarke). His crucial experiment consisted in letting various masses fall into a trough of clay (Fig. 7.1). He found that the depth of the impression in the clay depended on mv^2 rather than on mv, and he reputedly said '*Ah, c'est moi qui me suis trompé*'[6] ('Aha, it is I that is mistaken').

But 's Gravesande didn't leave it at that. He wanted to determine if mv^2 was an important measure in collisions as well as in free fall. He rotated his area of enquiry into the horizontal plane by letting spherical bodies suspended from a pendulum collide. Once again the motion was brought to a halt by clay or wax, and once again the depth of the impression was found to be proportional to mv^2.

's Gravesande also measured the mv before and after collision and found that the total mv was conserved, and this was true whether the

Fig. 7.1 Poleni's apparatus (similar to 's Gravesande's) for free fall into clay (photograph from the Boerhaave Museum, Leiden).

collision was elastic or inelastic. ('s Gravesande had an arrangement of ratchets and springs to measure this, in itself begging a lot of questions.) All in all, he concluded that *both* mv and mv^2 were important measures of 'force', but which measure came into play depended on the effects being investigated: mv was important when it came to a determination of the velocity after a collision; mv^2 was important when the 'force' was totally 'consumed' at the end of a fall or after a collision.

's Gravesande was on the right track, but it still wasn't evident to him that mv^2 and *mv together* are required for a full determination of the velocities after a collision. While, some 75 years earlier, Huygens had found that mv^2 was conserved in certain collisions (see Chapter 3), it was still hard to be sure what the significance of this was—was it just a number that happened to come out constant, or was there some new physical entity involved?

It appeared that *mv* had the edge over mv^2, as it was conserved in both elastic and inelastic collisions and was also conserved *at all instants* throughout the collision. Despite Leibniz's vaunting of the cosmic conservation of mv^2, in fact it wasn't conserved except at the start and end of the collision.

So where was the contest? It seemed as if the Newtonians were winning hands down (*mv* was conserved at all instants and in all types of collision), but the big trouble for Newton arose in a parallel metaphysical problem, the problem of hard-body collisions. In Query 31 in the *Opticks* (we have met it already in Chapter 3), Newton proposed that in the collision between two absolutely hard bodies approaching each other with equal and opposite speeds, as they couldn't penetrate each other, their motion simply stopped.

I believe that Newton's conclusion would have been just as startling and alien to the eighteenth-century philosophers as it is to us today. The trouble is that it simply never happens like that. Newton couldn't try the experiment with atoms (which by his definition were absolutely hard), but real composite bodies can be pretty close to being hard and impenetrable: yet the outcome of a head-on collision never even approximates to an abrupt cessation of motion, quite the reverse—the harder the bodies, the more strongly they rebound. (Edme Mariotte (1620?–84) tried using balls of ivory and of glass, as close to perfectly hard as was possible in those days.) Considering instead a collision between completely soft bodies, such as lumps of putty, then, again, the motion stops. For intermediate cases (the body has some elasticity), then motion is lost just to the extent that the body is inelastic. Newton gave no answer as to where the motion had gone. He simply said that if this absolute loss of motion is 'not to be allowed',[7] then we shall have to admit that bodies can never be truly impenetrable.

Johann Bernoulli didn't like the discontinuities that occurred in any collision let alone in Newton's query: while the *total* momentum was conserved, the momentum of an *individual* body could change abruptly, in magnitude and/or direction. 'Nature does not make jumps',[8] he complained.

Bernoulli and his son, Daniel (1700–82), also didn't like the implied net loss of 'cosmic motion' that resulted from Newton's query. Leibniz had said that the total 'effect' must neither exceed nor be less than the total 'cause', and both Bernoullis agreed with Leibniz and (remarkably) with each other on this. Thus the total cosmic motion had always to be conserved.

Why wasn't Newton more bothered by the counterintuitiveness of his thought experiment? Newton, like Leibniz and all the other natural philosophers, didn't sanction a net increase in motion coming from nowhere. His Third Law of Motion was just such as to prevent the 'absurdity' of such a perpetual motion (Chapter 3). However, Newton's thought experiment does not generate an increase but, rather, a net *decrease* in motion. This is something that *is* observed on the cosmic scale. As Newton observed, 'motion is much more apt to be lost than got, and is always upon the decay'.[9] Perhaps this is why Newton allowed his suggested thought experiment to remain.

The resolution came with compromises, eventually. Real bodies might be impenetrable, but they were compressible to a certain extent and approached absolute hardness only in the limit. (We can begin to appreciate the symbiotic relationship between concepts, experiment, and new mathematical ideas, such as 'in the limit'.) Johann Bernoulli helped physical intuition along by considering, hypothetically, a good analogy—an air-filled balloon. As the balloon was pumped up to a greater and greater pressure, it became harder and harder and also more and more elastic. At the limit of absolute hardness, the balloon became absolutely elastic instead of Newton's absolutely inelastic. Johann also considered the thought experiment of attaching perfectly elastic springs to perfectly hard bodies.

Jean le Rond d'Alembert (1717–83) made the parallel between the mathematical concept of a point and the abstract concept of hardness—neither was achievable in practice but both could be defined and then became usable concepts.

Finally, the problem of discontinuity could also be avoided by compromise. For a body that is not absolutely hard, there is some 'give' in it—and by stretching an instant of time to a small but finite interval, instantaneous changes (implying infinite forces) could be smoothed away (using the integral calculus).

So, what has happened to the contest? It seems as if *mv* has managed to recover after allowing for hardness to be an abstract idea, achievable only in the limit, and allowing for impact to occur over a short but finite time. The expression mv^2 also manages to survive provided that we understand 'conserved at all times' to really mean 'recoverable' at the end of the collision.

We still need to explain what happens in the case of two *soft* bodies colliding (putty, for example), and ask why we're not upset at the loss of motion in this case. Leibniz had the answer (and we have given it already in Chapter 3). He said that, while the bulk motion of the bodies did cease, no overall motion (*vis viva*) was lost, as it was preserved in the motion of the 'small parts'.

This consideration brings out a real difference between *mv* and mv^2. Leibniz had been referring to mv^2, but what about the momentum, *mv*—can it also be dispersed in the same way, by a transfer to the 'small parts'? The trouble, we now understand, is that bulk momentum has an overall direction and therefore can't all be transferred to random microscopic motions (these are in all directions). Leibniz suspected the trouble, as he claimed that only an undirected quantity could be of universal significance.

We now argue as follows. Consider the case of a soft body—say, a ball of putty—colliding with a wall or other immoveable object: where has the initial momentum of the putty gone? We can answer in two ways. First, we say that the wall is not truly immoveable but does, in fact, move a tiny bit in accordance with Newton's Third Law of Motion. As it has an almost infinite mass in comparison with the putty, the wall moves infinitely slowly: we can't see the motion of the wall, but it's enough to guarantee conservation of momentum (yet while introducing only a negligible amount of kinetic energy so that conservation of energy is not jeopardized!). Alternatively, we treat everything microscopically. The putty is deformed and heats up slightly, the wall also heats up and there may be a shock wave through it. All in all, the bulk motion of the putty is 'randomized' away.

A lot of this feels like arguing after the fact. It's no wonder that the philosophers of the day struggled with the concepts of *vis viva*, force, and hardness: the ideas were difficult, the mathematics was new, and the phenomena were detailed and various. Physical understanding and enlightenment (this was The Enlightenment, after all) would only come when experiment, mathematics, and ideas all moved forward together and when many different scenarios were examined.

So far, we have only looked at the case of collisions. Despite the high aims of the mechanical philosophy, this was perhaps not the best test case for a resolution of the controversy between *mv* and mv^2. In a collision the 'happening' is over so quickly that there's no time to witness the accelerations and decelerations (we shall find later, in the kinetic theory, that the collision-interaction itself is mysteriously idealized away). Besides which, Leibniz made truly cosmic claims for the applicability of *vis viva*. It was therefore necessary to examine as many different settings as possible in addition to the case of collisions.

Johann Bernoulli, despite being an ardent supporter of *vis viva*, made a pertinent criticism of Leibniz's use of the formulation mv^2. He said that gravity at the Earth's surface was just one example of 'activity'—what justification did Leibniz have in extending the validity of the formula, mv^2, derived from the specific case of free fall on Earth, to *all* phenomena? Leibniz had no good answer to this.

Turning our attention now to mathematics, we find a surprising thing—geometers (applied mathematicians) had been using Newton's definition of force (as in $F = ma$) for years. (The geometers included Varignon, the Bernoullis, Hermann, Euler, d'Alembert, and Clairaut—we shall meet them again in Parts II and IV.) More surprising, no one had attributed this relationship to Newton—least of all Newton himself (he attributed it to Galileo!). More surprising still, both *mv* and mv^2 were regularly coming out of the mathematics. Clearly, the geometers and the natural philosophers weren't talking to each other, even when these roles were combined into one person.

Newton had kept strangely quiet during the *vis viva* controversy—mind you, he was an old man by this time (he died in 1727 at the age of 84). However, he did pipe up at one point when he thought he had *vis viva* in trouble. When the force is constant (as in free fall), the acceleration is uniform and so equal amounts of *mv* are generated in equal intervals of time. However, in the case of *vis viva*, more is generated when the body's speed increases from 10 to 15 feet per second, say, than when its speed increases from 5 to 10 feet per second. Newton argued that this could only occur if the body's weight increased with time. This was clearly absurd, thought Newton, and so *vis viva* was evidently not the crucial determinant of 'motion'.

This seemingly paradoxical result occurred because Newton was looking at the effects following on from a compounding of force (Newton's F) through time. By looking instead at the compounding of force through *space*, a dependence on mv^2 rather than on *mv* appeared.

Newton himself had discovered this some 30 years earlier; he had effectively found that the area under the force versus distance curve was proportional to the square of the velocity.[10] However, the quantity mv^2 that arose out of his calculations held little significance for him—so little, it seems, that he promptly forgot all about it.

Eventually, by the middle of the eighteenth century, came the realization that a summation (strictly, an integration) of the force (Newton's F) for successive intervals of *time* generated a given change in mv, while a summation of the force for successive intervals of *distance* ($\mathrm{d}s$) generated a given change in mv^2:

$$\int F\mathrm{d}t = mv \tag{7.1}$$

$$\int F\mathrm{d}s = \tfrac{1}{2}mv^2 \tag{7.2}$$

At last we can see the factor of a ½ appearing before the quantity mv^2, turning it from *vis viva* into the familiar expression for kinetic energy, $\frac{1}{2}mv^2$. (In the mid-eighteenth century, units and constants hadn't yet settled down to common convention and dimensional analysis was still to be discovered by Fourier. A mere factor of ½ was not going to rumble Leibniz's new concept of *vis viva*.)

Once again, we might think that all the natural philosophers would say, well that's that then, the controversy is officially over—clearly *both* mv and $\frac{1}{2}mv^2$ are important measures. Some philosophers ('s Gravesande, d'Alembert, and others) did say just that, but most carried on just as partisan as ever. There were even some staunch Newtonians (such as George Atwood (1745–1807), well into the second half of the eighteenth century) who adhered strictly to Newton's Laws of Motion and rejected both mv and $\frac{1}{2}mv^2$. Just as happens today, it was one thing to do the mathematics and another thing to take on board the full meaning and implications of that mathematics.

Part of the difficulty lay in the fact that it was hard for the contemporary philosophers to accept the physical significance of an integration of force through space. Ironically, 100 years earlier, Galileo had struggled with the idea of introducing *time*. Now, in the mid-eighteenth century, no one wanted to consider anything but time. We have already seen how Newton tried to derail the *vis viva* concept by noting that it didn't increase uniformly with time. For Leibniz as well (a rare point of agreement between them), time was the crucial variable—in fact, Leibniz's whole philosophy demanded it. Everything from monads through to macroscopic bodies followed its destiny or world-line through time.

That any evolutionary change could come about by a mere change in position was ridiculous.

But, as we shall find out in Part II, this is exactly what can and does happen in nature.

In the meantime, what about the controversy? Which concept had won, mv or $\frac{1}{2}mv^2$? The curious thing is that the whole controversy, which had raged for over half a century, simply went off the boil. There were a number of reasons for this.

One is that the concept of the 'force of a body in motion' was fatally flawed. Galileo had shown that the motion of an isolated body can't be determined absolutely; and Newton showed in his First Law of Motion that a body in (uniform) motion needs no force to keep it moving (so, for example, the *Starship Enterprise* needs no fuel just to keep going). From our vantage point of over 200 years into the future, we can see that the controversy was really about energy—were Newton's concepts of force and momentum enough to describe the phenomena or was some other ingredient necessary? Atwood insisted that $\frac{1}{2}mv^2$ was merely a shorthand—Newton's Laws were complete and sufficient in themselves. But the quantity $\frac{1}{2}mv^2$ was cropping up in too many places to be dismissed in this way.

There are two ways in which motion can stop being relative and start being absolute. The motion can be altered and then the *change* relative to the former motion can be determined absolutely; the motion can be defined relative to other bodies *within a system*. These two ways are epitomized in the Newtonian 'force view' and the emerging 'energy view'. In the Newtonian view, an isolated body is the centre of attention and we track its twists and turns, its accelerations and decelerations, as it moves along its path and encounters the slings and arrows (the external forces) of fortune. In the energy view to come, a system is examined in its entirety and we look at the interplay between the various bodies within this system and within a certain time interval or cycle.

Paradoxically, the controversy waned just as the energy view was in the ascendant. Mechanics began to take a new turn away from an individual body and towards the system. The new systems view demanded its own new calculus (the variational mechanics of Euler and Lagrange; see Part IV), but this was so mathematical, abstract, and divorced from physical concepts that most geometers, engineers, and natural philosophers couldn't keep up.

The final reason why *vis viva* went off the boil is that it simply didn't answer to all that Leibniz had asked of it. Leibniz had shown great daring and intuition in inventing the concept—he had, in effect, discovered kinetic energy. But he wanted *vis viva* to play an even larger role, to account for all 'activity' of whatever kind and whether cosmic or local. It was to explain Leibniz's metaphysic that the total cause is equal to the total effect, whatever those causes and effects might be. But it turned out that the expression $\frac{1}{2}mv^2$ (kinetic energy) wasn't conserved except in the rather special case of elastic collisions. It had still to be realized that, yes, *vis viva* was a block of energy (when scaled by ½), but it was only conserved as mechanical energy when combined with another block of energy—potential energy.

Part II
Potential Energy

Introduction

It is potential energy that is the missing partner to kinetic energy. In a mechanical system (defined, it must be admitted, in a begging-the-question way as one from which there are no energy losses), it is only the combination of potential and kinetic energy, *taken together*, that is conserved. When kinetic energy is generated, it is at the expense of potential energy; when kinetic energy is consumed, potential energy is increased in equal measure. But what is potential energy? Whereas kinetic energy had a specific formulation and a discoverer (Leibniz), we shall see that the formulation and understanding of potential energy just sort of slipped into use—a gradual process taking 100 years or so.

There were three routes along which ideas about potential energy began to develop. The first was the intuitive understanding that the live force doesn't just disappear or appear from nowhere; there has to be some quantity that represents live-force-in-waiting—but this intuition didn't immediately yield a formula for potential energy. The next route was the solution of special problems by geometers. They used Newton's Second Law to solve problems such as the orbits of planets and the motion of vibrating strings. Finally, the third route was via the engineers—they began to define a quantity that they found useful when making comparisons between the performance of different machines. This was the quantity called 'work'.

Work

We have emphasized how new concepts in physics are only forged when experiment, mathematics, and the ideas themselves all move forward together. Now we see yet another factor that is important—technology. We shall examine this last case first.

For the most part, the conservation of momentum was a strangely unhelpful rule for the engineer—although the conservation of total *mv* *was* demonstrated in the ballistic pendulum invented by Benjamin Robins (1707–51). Even stranger, Leibniz's *vis viva* also had only a limited utility. A much more useful measure began to appear—that quantity called 'work'. This the engineers took to be the 'force' (tacitly, Newton's *F*) multiplied by the distance through which the force acts. We have already seen (Chapter 4) how Amontons, as early as 1699, had said that the 'labour of a horse' was the force exerted by the horse multiplied by the speed of the horse at its point of application (strictly speaking, this is a measure of the *rate* of doing work).

The engineers' incentive was the need to make comparisons between a growing variety of machines: the water-wheel and the windmill, the Archimedean screw and the hand pump, the horse-drawn plough, the labourer with wheelbarrow or shovel, and the increasing number of steam-driven water pumps. The variety was important. That the concept of work, initially derived for the engineers' expediency, could be used in all these cases meant that here was something fundamental and universal. Mathematically, it can be seen that 'work' is none other than the geometers' integration of force over distance (Equation 7.2). As this is equal to *vis viva*, then work is equal to *vis viva*. But the geometers and the engineers moved in different worlds and so the connection between work and *vis viva* took time to be noticed.

However, in the special case of gravity, the connection between 'work' and *vis viva* was pretty well axiomatic (weight × height is proportional to mv^2; Leibniz, see Chapter 3). Furthermore, the weight-raising machine maintained a special hold on the engineers' attention during the eighteenth century, as the most ubiquitous machine was the water pump for draining mines (around 20 times more water came out of English mines than ore).

The lever was the quintessential weight-raising machine but already by the sixteenth century Stevin had shown that all the simple machines (the pulley, wedge, screw, inclined plane, etc.) were really equivalent to the lever. It was then a gradual but inevitable step to generalize the definition of work from the lever to all simple machines, and then to all real machines, simple or not and weight-raising or not. Amontons, again, was prescient in applying the concept of work to his 'fire-engine' (see Chapter 4). This was an idea that was years ahead of its time.

One other philosopher who was years ahead of his time was the French engineer-scientist, Antoine Parent (1666–1716). In 1704, he formulated his 'Theory of the greatest possible perfection of machines'.[11] He examined water-wheels and made a crucial connection between the *vis viva* of a flowing stream and the 'work' that a water-wheel could do. He was the first to ask what was the maximum power ('*effet*') that could be extracted from the stream, and he answered: it is equal to the mass of water times the height reached when the stream is directed upwards to make a fountain.

Parent's next question was: what is the connection between this '*effet*' and the work of a stream *on the flat* and turning the blades of the wheel? He realized that all wheels worked in the same way, whether exploiting a fall of water (as in the overshot wheel) or a moving stream (the undershot wheel). Impressively, he used the new calculus, but he made certain unwarranted assumptions and found that the maximum theoretical efficiency of any water-wheel was 4/27, which was soon found to be a serious underestimate.

Another French engineer, Antoine Déparcieux, used another original and impressive line of argument (in 1752). He said that the maximum efficiency had to be '1', as an overshot wheel could drive another wheel in reverse (using scooped buckets). Assuming mechanical perfection, then only a very small excess of water would be needed for the forward wheel to drive the reversed wheel.

The English engineer John Smeaton, in 1759, showed that overshot wheels could extract around twice as much work as undershot wheels (from the same stream). The Newtonian 'force-mechanics' could give no explanation for this difference: total momentum was conserved whichever type of wheel was used. The continental *vis viva* mechanics *could* provide an explanation. The undershot wheel was situated on a fast-flowing stream and there was considerable turbulence and 'shock' to the wheel-blades; the overshot wheel was fed from slow-moving water from a millpond and there was therefore less loss of *vis viva*, less waste of potential work.

Thus a new concern, an *engineers'* concern—the efficiency of machines—led to new insights.

There were those natural philosophers who weren't that impressed by all this arguing from real machines. D'Alembert, as we have noted, was one who tried to axiomatize mechanics from pure reason. Even Newton, while a gifted experimenter, was curiously silent on the new 'power' technology of water-wheels and heat-engines.

But the links between work and *vis viva* slowly became more and more compelling, although they were not put into a theoretical framework until the work of Lazare Carnot (1753–1823) in the second half of the eighteenth century. This new concept, work, justified its existence by its utility to the engineers and by its universality, being applicable whatever the force and whatever the machine.

Origins

Let's now turn back to the first route to potential energy—the philosophers' intuitive understanding of potential energy, negatively, as a shortfall in kinetic energy. They understood potential energy as both a reservoir and a sink of kinetic energy.

Leibniz had the germ of an idea of potential energy. He called it *vis mortua* (dead force) and considered that it existed in cases of equilibrium where motion had not already started—it was a 'solicitation to motion'.[12] Once new motion had been generated then 'the force is living and arises from an infinite number of continuous impressions of dead force'.[13]

But Leibniz did not consider that there was anything symmetrical about these two kinds of 'force'; there was no double-act, no constant exchange or interplay between potential and kinetic energy. In fact *vis mortua* and *vis viva* were, for Leibniz, two very different kinds of entity:

> living force is related to bare solicitation as the infinite to the finite, or as lines to their elements in my differential calculus.[14]

Johann Bernoulli, who, as we've seen in Part I, was a vociferous supporter of Leibniz's *vis viva*, recognized many examples of stored 'live force', for example, in a stretched spring and a raised weight. Johann also moved a step closer to perceiving an equivalence between the dead and the live forces. He stated that *vis mortua* was consumed whenever *vis viva* was generated and that the reverse was also true. This had to be so in order to guarantee that the cause was equal to the effect and vice versa.

Nevertheless, Johann still agreed with Leibniz that the relation of *vis mortua* to *vis viva* was like 'a line to a surface' or 'a surface to a volume'.

We now agree with Johann Bernoulli that 'live force' may be stored as in a longbow, a pendulum, a squashed spring, a boulder, and so on. But there's more to potential energy than mere potential. The common feature of all these examples is that they all exhibit a dependence on *position* (the bowstring must be pulled back a few inches, the bob raised a certain height—and also moved laterally—the spring squashed in or stretched out, and the boulder moved up a mountain from ground level). This last example brings out another telling feature—it is only the *relative position* that is important. The same is, in fact, true for all the other examples—the dependence is always on relative rather than on absolute position.

The concept of potential energy did not win a quick or an easy acceptance. There was a tacit use of it for years before it was brought out into the open—and named (by Daniel Bernoulli; see Part III). Something 'actual', such as motion, is easier to understand than something latent. Also, *vis viva* and force are more intuitively obvious concepts as they both relate to an individual body. Potential energy relates instead to the relative positions of bodies or parts of a body *within a system*—it should more properly be called the energy of relative position or of configuration.

As the 'geometers' began to deal with special problems involving more and more complicated scenarios (systems), the concept of potential energy began to emerge. The solutions to these problems would forge the new discipline of mechanics which would in retrospect be called Newtonian mechanics (although the practitioners—at least, those outside Britain—did not feel that they were contributing to a Newtonian legacy).

Clairaut

A problem that led to one of the most entertaining episodes in natural philosophy was that concerned with the 'figure' (shape) of the Earth. Newton predicted that it was fatter towards the equator, whereas the French Royal astronomer, Jacques Cassini (1677–1756), thought the Earth was elongated along its axis.

The French Royal Academy of Sciences arranged and funded two expeditions to test these competing predictions. The philosophers had to measure the length of a one-degree change in latitude along a given line of longitude. One team would visit Quito on the equator, the other team would go to Lapland in the polar region. It is with this second group that our story continues.

The philosopher and mathematician Pierre-Louis Moreau de Maupertuis, the youthful Alexis-Claude Clairaut, the Swede Anders Celsers (of the Celsius scale), and others left Paris for Lapland in 1736. They had many adventures, including being welcomed by the King of Sweden, being ice-bound, having their signalling fires removed for firewood, enduring plagues of gnats, and so on. The exigencies of this undertaking can hardly be exaggerated (the expedition took almost two years, while the team in South America took over ten years with some never returning). The eventual results supported Newton's theory. Voltaire was delighted and described Maupertuis as the 'flattener of the poles and the Cassinis'.[15] (He also lampooned Maupertuis for bringing back to Paris two attractive Lapland women.)

Now comes the bit of relevance to energy. Alexis-Claude Clairaut (1713–65) determined the equilibrium shape of a rotating, self-gravitating fluid mass (the Earth) in his book *Théorie de la figure de la terre*[16] in 1743. With a totally original argument, he considered hypothetical canals within the body of the Earth (Fig. 7.2) and saw that equilibrium would be maintained so long as no work was done (by a small test mass) in going from one end of the canal to the other. He realized that the work would indeed be zero when the canal end-points had the same

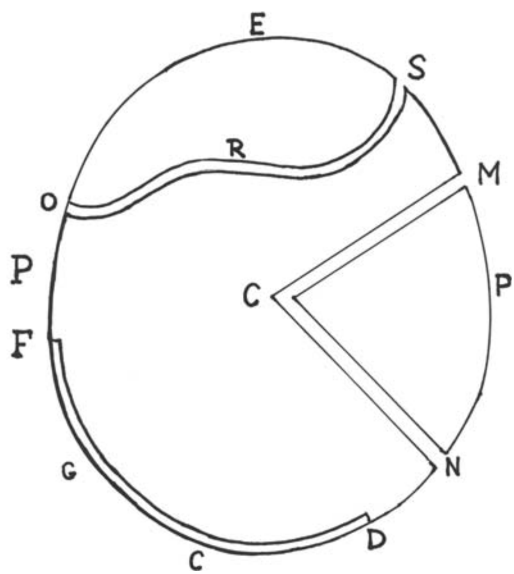

Fig. 7.2 Clairaut's 'canals', in his *Théorie de la figure de la terre*, 1743.

value of some function—the 'potential function'. This was the very start of potential function theory, continued by Pierre de Laplace at the end of the eighteenth century and formulated rigorously by George Green in 1828.

Let us now continue with the other geometers and special problems of this period.

Johann Bernoulli

Johann, as we have seen earlier, understood the idea of potential energy as stored 'live force'. He used the accepted 'rule of accelerative forces', $F = ma$ (of course, this had nothing to do with Newton as far as Johann was concerned) to solve various problems. For example, in 1727, he considered the vibrations of a string (such as a violin string) fixed at both ends. (The mathematician and musician, Brook Taylor—see Chapter 5—worked independently on the same problem.) Johann implicitly identified the product of the tension and extension of the string with potential energy and found that the maximum potential energy was equal to the maximum *vis viva* of the string. He also carried out experiments and found that Hooke's Law (Mariotte's Law in France), that the extension was proportional to the force, didn't always apply—he was using gut, which is anomalous in this regard.

Johann also had the idea of 'work'. However, he made no link between the engineers' concept of work and his own. This was because he used it to solve problems in statics rather than in power generation. In fact, he used the idea of work to solve the case of static equilibrium in a novel way, and in a way that was to have enormous ramifications for mechanics. This was Johann's Principle of Virtual Work and it was to be a founding principle of Lagrange's analytical mechanics. It was never published by Johann but was explained in a throw-away letter to Pierre Varignon in 1717. Fortunately, Varignon didn't throw it away. We shall discuss it further in Part IV.

Daniel Bernoulli

Daniel, like his father Johann, attached huge importance to the concept of live force and also understood potential energy as a store of live force. In fact, Daniel was the philosopher who first introduced the term 'potential'. But Daniel went further than his father in appreciating the equal status of potential and actual live force. For example, for Daniel, the fact that the centre of gravity of an isolated system maintains a constant height was explained by the '*equality between the actual descent and potential ascent*'.[17] This was preferred to those 'Paternal words', the '*conservation of live forces*'.[18] There are countless references to 'actual' and 'potential' live force in Daniel's book, *Hydrodynamica*,[19] published in 1738.

Daniel's later work in celestial mechanics showed that he made the connection between the geometers' equation, $\int F\mathrm{d}s = \frac{1}{2}mv^2$ (Equation 7.2) and the equality between potential and actual live force. This led Daniel on to making the association between $\int F\mathrm{d}s$ and potential 'energy'—a totally modern outlook.

Daniel was one of the first (*the* first?) to recognize the important link between the stability of certain mechanical systems and the fact that the potential energy was at a minimum in these cases. He passed this intuition on to Euler and it helped the latter take the first steps towards the variational mechanics (see Parts III and IV).

Daniel Bernoulli was the only one of his contemporaries who saw kinetic and potential energy in a wider context, outside the confines of mechanics and in the real world of engines and machines. One can sometimes substitute the modern word 'energy' for the old term 'live force' in his work without disturbing the meaning. It is for this reason that we consider him to be a hero of energy in the eighteenth century. His life and work is reviewed in Part III below.

Euler

The leading applied mathematician of the eighteenth century—and, perhaps, of all time—Leonhard Euler (1707–83) was certainly the most prolific, publishing major books and an average of a paper a week for the whole of his working life. This was in the midst of the hubbub of family life (he had 13 children—eight of whom died in infancy—and 30 grandchildren). Overwork led him to lose the sight of one eye in 1735. In later life his other eye deteriorated and he was forced to do his calculations in large characters on a slate. Finally, in 1766, he lost his sight altogether, but his work continued unabated (he carried out the calculations in his head). His publications also continued unabated as he dictated them to his pupils and his children—in one case even his tailor was enlisted (the tailor wrote down Euler's *Elements of Algebra*).

In 1736, Euler published his *Mechanica, sive motus scientia analytice exposita*'.[20] This was a landmark in mechanics, flanked by Newton's *Principia* in 1687 and Lagrange's *Analytical Mechanics* in 1788. In it Euler put (Newtonian) mechanics on rigorous, analytical foundations: formulating Newton's Second Law as $F = ma$; defining angular momentum and introducing its conservation as a separate axiom in addition to the conservation of linear momentum; defining the point-mass, the rigid body, and the continuum; and identifying the rotation of a rigid body about three independent axes. (Incidentally, this last item led to an extension of the formulation of kinetic energy from $\frac{1}{2}mv^2$ for translations to $\frac{1}{2}I\omega^2$ for rotations.*)

In a number of his papers, Euler arrived at the formulae for kinetic and potential energy and saw that these were conserved quantities. Why do we not consider Euler also to be a hero of energy? The answer is that Euler did not base a whole physics on energy and, indeed, hardly ever used the term 'live force'. He was hostile to Leibnizian philosophy (as espoused by Leibniz's follower, Christian Wolff), especially the 'monadology' in which there was an active force within each monad. He explained in his *Letters to a German Princess*[21] that the force had always to be *external* to the body in question.

Euler's views on force were therefore more aligned with Newton than with the Leibnizians. He may also have had religious objections to the latter. He was a staunch Lutheran and believed in a God by revelation rather than in a God who had rational requirements, such as the principle of sufficient reason and so on. (Wolff retaliated by saying that 'Euler is a baby in everything but the integral calculus'.[22])

Boscovich

A Jesuit priest from Ragusa (present-day Dubrovnik) in Dalmatia, Roger Boscovich (1711–87) had some far-reaching and iconoclastic scientific views. In order to defend the principle of continuity and to solve the problems of hard bodies in collision, Boscovich developed a new

* I is the moment of inertia and ω is the angular speed.

theory of matter in which atoms were 'points of force'[23]—indivisible and without extension. Instead of forces operating by contact (as advocated by Euler), Boscovich admitted only action-at-a-distance. The points exerted forces on each other dependent only on their mass and their separation. At very short separations there was a strong repulsive force which increased to infinity as the points were brought closer together. This infinite repulsion prohibited the points from ever coming into contact. As the separation was increased, the force alternated between repulsion and attraction, and finally (for large separations) followed Newton's inverse square law of gravitational attraction.

Boscovich illustrated his theory by his famous curve of force against distance. The diagram looks like a modern potential energy diagram. Certainly, Boscovich knew that the area under the curve was a constant of the motion, but he didn't extract the concept of energy from it. In fact, he was antagonistic to the idea of *vis viva*, perhaps for the same reasons as Euler.

In defining force as something that acts at an abstract point in space, Boscovich's work was a precursor to modern field theory: apparently Michael Faraday (see Chapter 15) studied Boscovich with care.

Boscovich went on to write poetry and to annotate Benedict Stay's *Principia* in verse form. (This was a pastime of many natural philosophers in the eighteenth century. For example, Erasmus Darwin, grandfather of Charles, put forward a proto-theory of evolution in the poems 'The Temple of Nature' and 'The Loves of the Plants'.[24])

Laplace

Some 30 years later, in 1785, the mathematician Pierre Simeon Laplace (1749–1827) calculated the Newtonian gravitational attraction at any arbitrary point outside an extended spheroidal mass: for example, the attraction of the Earth on a cricket ball. His formulation became known as Laplace's equation—although it had already been discovered by Euler and by Lagrange (who will be introduced in Part III below).

For a cricket ball with (rectangular) position coordinates x, y, and z, the potential (V) was a function just of x, y, and z and, remarkably, didn't depend on the distribution of mass within the Earth. The equation was remarkable in another way—it turned out to be applicable to a large proportion of the problems in physics. In it, the dependence of potential energy on configuration was more complicated and abstract than the simple cases of the extension of a spring or the release-height of a pendulum bob. However, it wasn't until well into the nineteenth century that Laplace's equation was incorporated into the body of physics and could, in retrospect, be seen as a bridge between the work of Clairaut in 1743 and Green in 1828.

Summary

By the second half of the eighteenth century, the ideas of potential and kinetic energy were beginning to emerge. Kinetic energy was the energy of motion, now generalized to include rotations as well as translations. Potential energy was more than just kinetic-energy-in-waiting, it was the energy of configuration.

Potential energy and work were eventually seen to be one and the same (an integration of the force with respect to distance). However, the terms 'work' and 'potential energy' grew to have a subtly different usage: work was the energy used up or generated by a force acting along a specific path in space, whereas potential energy could be defined at a specific point in space (albeit relative to some given reference location).

The concept of potential energy thus carries over into the modern field theory where various properties (e.g. the height and direction of a blade of grass) can be defined at every point within the field. On the other hand, the fact that work is defined *along a path* is a feature that will have especial relevance to the next advance in mechanics, the variational mechanics (Part IV).

The curious thing is that while mv^2 and force, F, were discovered by specific people—Leibniz and Newton, respectively—and arrived with much fanfare and controversy, potential energy just sort of slipped in uninvited, like the thirteenth fairy at the party.

Part III
Daniel Bernoulli, Unsung Hero of Energy

We have said how the controversy over *vis viva* went off the boil in the second half of the eighteenth century. If there was one philosopher who kept the concept at least simmering through the middle years of the century, it was Daniel Bernoulli (we have met him before, both here and in Chapter 6). In his great work, *Hydrodynamica*, he stated the conservation of living force as his most important principle:

> The primary one [principle] is the conservation of living forces, or, as I say, the equality between *actual descent* and *potential ascent*.[25]

In one fell swoop, Daniel Bernoulli gave primacy to energy conservation and introduced, for the first time, the term 'potential'. This was a progression from his father's views and was a step towards seeing kinetic and potential energy in more equal terms.

Daniel, like his father, had started his studies in medicine rather than in physics, and at the end of his career he held professorships in both physiology and physics. This gave him a wider stage on which to see the applicability of the concept of living force. His doctorate on respiration and an early paper on the effect of motion on the muscles utilized the concept, and in a talk on the work-rate of the heart he made explicit reference to the conservation of living force.

At the same time, Daniel had no woolly ideas about some vague, vital spirit permeating everything. Always, his use of the principle was strictly quantitative. We have seen from Feynman's parable of Dennis' blocks and from Galileo's exhortation to describe the world mathematically that new concepts must be accompanied by precise formulae. Daniel Bernoulli certainly upheld this requirement. Time and time again in his book *Hydrodynamica*, he uses wordy constructions in order to ensure that concepts such as force, live force, and potential live force are quantitatively defined. For example:

> the force of a fluid dashing against a perpendicular plane at a given velocity is equal to the weight of the cylinder of fluid erected above that plane, of which the altitude is such that from it something moveable, by falling freely from rest, would acquire the velocity of the fluid.[26]

Hydrodynamica shows that in two other respects (as well as his advocacy of the energy principle) Daniel was different from his contemporaries: he believed in checking his theories out against experiment; and he was interested in machines—the limits to their perfectibility and their utility to mankind. Bernoulli was, in other words, more of a *physicist*, whereas d'Alembert, Clairaut, Euler, and Lagrange were more akin to applied mathematicians—to use the modern terminology.

In Chapter IX of *Hydrodynamica*, Bernoulli introduced the word 'potential' again—this time in the context of a hydraulic machine: he defined the *absolute potential* as given by 'a weight, an activated pressure, or other so-called dead force…multiplied by the distance which the same moves through'.[27]

He then went on to make quantitative comparisons between the *absolute potential* of the machine and the 'work endured by day labourers elevating water'.[28] It is a bit startling to read that the work 'is not to be interpreted in a physiological, but in a moral sense', but then he continues: 'morally I estimate neither more nor less the work of a man who exerts at some velocity a double effort than that of one who in the same effort doubles the velocity, because certainly either one achieves the same effect, although it may happen that the work of the one, despite being no less strong than the other, is very much greater in a physiological sense'.

Bernoulli thus carried on by a thread the tradition that examines the *efficiency* of machines—from Amontons and Parent before him and to the Carnots, father and son, around the turn of the eighteenth to nineteenth century (see Chapters 8 and 12). Bernoulli concluded (with regard to a hydraulic machine) that:

> there is some definite termination of perfection beyond which one may not be able to proceed.

and

> With the same absolute potential existing, I say that all machines which suffer no friction and generate no motions useless to the proposed end maintain the same effect, and that one [design] is therefore not to be preferred to the other.[29]

Such observations were to have revolutionary repercussions for thermodynamics in the nineteenth century.

In the last chapter of *Hydrodynamica*, Bernoulli put forward an innovative scheme for driving ships (what we would now refer to as jet propulsion):

> I do not see what would hinder very large ships from being moved without sails and oars by this [the following] method: the water is elevated continually to a height and then flows out through orifices in the lowest part of the ship, it being arranged so that the direction of the water flowing out faces towards the stern. But lest someone at the very outset laugh at this opinion as being too absurd, it will be to our purpose to investigate this argument more accurately and to submit it to calculation.[30]

The most original part of the book from the perspective of the developing concept of energy occurs in the chapter on elastic fluids (gases). We have already seen (in Chapter 5) how Bernoulli proposed a kinetic theory of gases. His vision of a gas as made up of an almost infinite swarm of minute particles in ceaseless motion foreshadows our modern view (and contrasts with the contemporary 'passion-fruit jelly' model). He accounts for the pressure of a gas as being due to collisions of the gas particles with the container walls, and finds both that Boyle's (Mariotte's) Law should be followed and that the pressure is proportional to the square of the speed of these gas particles. A temperature increase would—Bernoulli says—lead to an increase in the particle speed and this in turn would lead to an increase in pressure. These ideas are almost identical to our modern views (the difference being that Bernoulli makes no distinction between speed and average speed).

Bernoulli continued with a further feat—he managed to derive a quantitative expression for the 'live force' contained in the gas. How he did this was as follows. He considered the gas cylinder again, but this time the piston was imagined loaded with an extra weight that was driving it down. He then applied Newton's Second Law of Motion to the falling piston and so found its speed and its kinetic energy ('actual live force'). He then equated this to the 'potential live force' of the piston (its weight times the fall-height). But the two were not equal, and Daniel Bernoulli drew the correct conclusion—the discrepancy was due to the extra live force needed to compress the gas (what we now refer to as the work done by the piston on the gas):

> I say, therefore, that *air occupying the space [volume] A cannot be condensed into the [volume] space A – h unless a live force is applied which is generated by the descent of the weight [pressure], p, through the height, A ln(A/(A – h)), however that compression may have been achieved; but it can be done in an infinite number of ways.*[31]

(The initial cylinder height is A and it is reduced to $(A - h)$ after the piston has fallen through a height h.) Crucially, Bernoulli now had an expression for this live force: pressure, p, times $A \ln(A/(A - h))$.

Granted that $A \ln(A/(A - h))$ is effectively the same thing as a change in volume, $\mathrm{d}V$, we end up with the formula $p\,\mathrm{d}V$ for the work done by a piston on a gas. The expression $p\,\mathrm{d}V$, another block of energy, was to become an iconic component of classical thermodynamics. However, this part of Bernoulli's work was never examined, and the expression was lost and then rediscovered over 50 years later (in 1792, by Davies

Gilbert; see Chapter 8), after Watt's invention of the steam engine (Chapter 8).

Of course, Daniel Bernoulli was not able to see into the future, but he did appreciate the value of his formulation:

> this is an argument worthy of attention, since to this are reduced the measures of the forces for driving machines by air or fire or other motive forces of this kind, of which perhaps several new ones could be developed, but not without considerable practical mechanical improvement and perfection...[32]
>
> since it happens in many ways that air is compressed not by force but by nature... [then] there is certainly hope that from natural occurrences of this kind great advances can be devised for driving machines, just as Mr. Amontons once showed a method of driving machines by means of fire...
>
> I am convinced that if all the live force which is latent in a cubic foot of coal and is brought out of the latter by combustion were usefully applied for driving a machine, more could thence be gained than from a day's labour of eight or ten men.[33]

Bernoulli immediately put his new formula to the test. Stephen Hales, in his *Vegetable Staticks* (see Chapter 5), had found that 'Half a cubick inch, or 158 grains of Newcastle coal, yielded 180 cubick inches of air.'[34] Therefore (Bernoulli calculated), a cubic foot of coal would yield 360 cubic feet of air and this would correspond to the 'falling of a weight of 3,938,000 pounds from a height of one foot.'[35]

Of steam, Bernoulli wrote:

> water reduced to vapour by means of fire possesses incredible force; the most ingenious machine so far, which delivers water to a whole town by this principle of motion, is in London.[36]

This was very likely Newcomen's revolutionary new engine (see Chapter 5).

Finally, Bernoulli wrote of 'the astounding effect which can be expected from gunpowder'. His calculations showed him that:

> in theory a machine is given by means of which one cubic foot of gunpowder can elevate 183,913,864 pounds to a height of one foot, which work, I would believe, not even 100 very strong men can perform within one day's span, whatever machine they use.[37]

While today we no longer adhere to Bernoulli's view that coal and gunpowder contain 'air' compressed to a very high degree, we see that Daniel Bernoulli's analysis demonstrates a highly original interpretation of the concepts of actual and potential live forces. He brings them ever closer to our modern conception of energy. However, his work was picked up by only a few members of the Swiss school (for example, Pierre Prévost; see Chapter 6) and then only to the extent of the kinetic theory itself and not these later sections of Daniel's Chapter X).

The energy principle was so important to Daniel that it permeated almost all his subsequent work after *Hydrodynamica*. For example, he considered the case of a deformed elastic band (think of a ruler flexed into various shapes). In a letter to Euler written on 7 March 1739, he proposed that when the potential energy of the band is at a minimum, then the equation of the elastica (the shape of the curve) would follow:

> I have today a quantity of thoughts on elastic bands... I think that an elastic band which takes on of itself a certain curvature will bend in such a way that the live force will be a minimum, since otherwise the band would move of itself. I plan to develop this idea further in a paper; but meanwhile I should like to know your opinion on this hypothesis.[38]

Euler replied (on 5 May 1739):

> That the elastic curve must have a maximum or minimum property I do not doubt at all... but what sort of expression should be a maximum was obscure to me at first; but now I see well that this must be the quantity of potential forces which lie in the bendings: but how this quantity must be determined I am eager to learn from the piece which your Worship has promised.[39]

But Euler still had years to wait. In 1742, Bernoulli wrote:

> My thoughts on the shapes of elastic bands, which I wrote on paper only higgledy-piggledy and long ago at that, I have not yet been able to set in order...[40]

and finally, in 1743:

> May your Worship reflect a little whether one could not deduce the curvature... directly from the principles of mechanics... I express the potential live force of the curved band by $\int ds/r^2$... Since no one has perfected the isoperimetric [variational] method as much as you [Euler], you will easily solve this problem of rendering $\int ds/r^2$ a minimum.[41]

Daniel's idea that the stable state of the band occurred for a minimum in the potential energy was crucial in the emerging variational mechanics (to be covered in Part IV below). As Euler was to write:

> although the curved shape assumed by an elastic band has long been known, nevertheless the investigation of that curve by the method of maxima and minima [could not be carried out until the] most perspicacious Daniel Bernoulli

pointed out to me that the entire force stored in the curved elastic band may be expressed by a certain formula, which he calls the *potential force*, and that this expression must be a minimum in the elastic curve.[42]

(In other words, Daniel needed help with the maths and Euler with the physics.)

There were other cases where Daniel employed a minimum principle. For example, already over 70 years old, he analysed the motion produced in a bar that had been struck at its mid-point. By assuming that the shape adopted was such as to minimize the kinetic energy, he was able to predict the subsequent motion of the bar. (This became, in the twentieth century, the Rayleigh–Ritz Principle.)

While Daniel's father, Johann Bernoulli, was deeply antagonistic to all theories Newtonian, Daniel actively supported Newton and promoted his work on the continent (he referred to Newton as the 'crown prince'[43] of physics). In fact, it was Daniel's open-mindedness and his ability to merge the British Newtonian and continental Leibnizian traditions that really opened the door to the new energy physics of the nineteenth century.

In a paper in 1738, Daniel used Newton's inverse square law of gravitational attraction to solve the problem of the motion of the Moon in a Sun–Earth–Moon system. However, Daniel's approach was energy-based rather than force-based—completely novel in this celestial setting. He calculated the *vis viva* of the Moon and the work done by it as it was brought to its orbit from infinity. This was only slightly more outrageous an approach in the mid-eighteenth century than it would be today. (Who now would calculate the kinetic energy of the Moon?) Euler's evaluation of this paper showed its importance:

I was pleased no end about His Honourable etc [Daniel Bernoulli's] '*Dissertation de Principio virium vivarium*', (firstly) for the idea itself to apply the '*Principium Conservationis Virium Vivarum*' to this material, as well as for the usefulness and incredible advantage which one gains by it, to find out the movement of the Moon, which otherwise, and only approximately so, must be derived by the most intricate equations.[44]

Daniel planned to extend this work with a more general treatment considering several force centres. This he carried out 12 years later in 1750. He considered a system of two attracting bodies, then three, and finally the general case of *n* attracting bodies. He implicitly assumed central forces—forces dependent only on the separation between the body and the force centre.

He found that as the bodies travelled between given initial and final positions, the increase in the total live force was dependent only on these terminal positions and not on the route taken. He had effectively calculated the potential energy of a system of *n* interacting bodies and found that the potential energy plus the kinetic energy summed to a constant value (the total energy, as we would say now). But Daniel preferred to interpret the results in terms of the route-independence of any changes in *vis viva* between two end points.

Bernoulli had considered only two types of central force—a constant (parallel) force field and Newton's inverse-square attraction (the former case could be taken as central, with a force centre removed infinitely far away). He stated his concern that the applicability of inverse-square attractions in non-terrestrial regions had not yet been demonstrated by observations. However, he understood that his energy principle was of general validity and would still follow as long as the forces were central. It is noteworthy that while developing these ideas of kinetic, potential, and total mechanical energy, Bernoulli still adhered to Newton's concept of force, considering it as real and not just a convenient shorthand.

Why did Bernoulli's work go unnoticed? Why is he unsung, even to this day?

We have seen how Bernoulli was ahead of his time with regard to the kinetic theory of gases (Chapter 5), calculating the work done by a gas, applying a minimum principle to mechanical problems, the general use of the 'energy principle', and the premonition of stored live force as a fuel or source of power. He was an original and creative thinker in many other areas as well. In the foundations of mechanics, he examined the parallelogram law for adding forces—was this law empirically or necessarily true? He also generalized collisions between bodies to include the case of angular momentum transfer as well as the transfer of linear momentum. In analysis, he described the vibrations of a stretched string, fixed at both ends, as a superposition of an infinite number of simple harmonic terms. He also tackled some perenially hard problems such as analysing the tides and analysing rolling and sliding friction. His contributions to experiment and engineering are too numerous to mention here.

He is most famous today for the 'Bernoulli effect'—the continuity condition requiring that fluid flowing around a curve speeds up relative to fluid flowing via a more direct route. (The faster-flowing fluid exerts less pressure to the sides and this results in 'lift'—for example, for air flowing around an aircraft wing.)

Outside of mathematical physics, Daniel Bernoulli was a leader in the area of probability and statistics. For example, he calculated the change in life expectancy of a population before and after inoculation against smallpox ('variolation')—this work influenced Catherine the Great.

However, in the area of mechanics, Bernoulli was overshadowed by Euler (see Parts II and IV), the greatest applied mathematician of the age and possibly of all time. Also, Bernoulli was by-passed by a move towards a more abstract mechanics led by d'Alembert, Euler, and Lagrange (to be covered below). This wasn't Bernoulli's style—he liked to keep a grip on the physical implications of his theories and also to carry out experimental investigations to get a physical feel for things. In short, Daniel Bernoulli was that rare and unfashionable thing at the time, a true physicist.

Of course, Daniel Bernoulli may just not have had the right sort of personality to promote his ideas. We know very little about his private life, but he seems to have been more affable than his father or uncle. These two were often bitter adversaries, whereas Daniel was very close to his brothers (also mathematicians), even though they were more favoured by his father. (This is evidenced in Daniel's letters, where we find, for example, that he was very upset by the sudden death of his older brother Nikolaus II in St Petersburg.) Daniel corresponded with all the leading mathematicians of the age and was a close friend to some (Euler and Clairaut).

On finding that his father had pre-dated the *Hydraulica* and used many of Daniel's results without acknowledgement, Daniel wrote in complaint to Euler:

> Of my entire *Hydrodynamica*, of which indeed I in truth need not credit one iota to my father, I am robbed all of a sudden, and therefore in one hour I lose the fruits of a work of ten years. All propositions are taken from my *Hydrodynamica*; nevertheless my father calls his writings *Hydraulica, now first discovered, anno 1732*, since my *Hydrodynamica* was printed only in 1738 . . . In the beginning it seemed almost unbearable to me; but finally I took everything with resignation; yet I also developed disgust and contempt for my previous studies, so that I would rather have learned the shoemaker's trade than mathematics. Also, I have no longer been able to persuade myself since then to work out anything mathematical. My entire remaining pleasure is to work some projects on the blackboard now and then for future oblivion.[45]

Fortunately for physics, Daniel Bernoulli did not carry out this threat, although it is true that he never again reached such a peak in creativity. His book *Hydrodynamica* was written while he was at St Petersburg, between 1725 and 1733 (he was invited there when Peter the Great founded the Russian Academy of Science). Daniel encouraged Euler to join him at St Petersburg (Euler's relative Jakob Hermann was also there at this time). But the Russian climate didn't agree with Daniel and he moved back to Basle in 1733, to become professor of botany and anatomy. Daniel's father held the mathematics chair at Basle but when he died in 1748 the chair skipped Daniel and went to his younger brother Johann II. However, two years later Daniel acquired the physics chair, which he held until his death in 1782 (he also managed to swap botany for physiology). His lectures were accompanied by entertaining experimental demonstrations. Daniel never married. His nephews helped him with lecturing duties in the last few years of his life.

Part IV
Variational Mechanics, a Very Principled Tale

The Principle of Least Action

We now come to a new departure for mechanics and for the whole of physics. A new metaphysical principle will be brought in and a new scalar quantity defined. Up to now we have stressed the conserved quantities of the mechanical system—the energy, the momentum, the direction and speed of a body subject to no forces, and so on. Now (in the mid-eighteenth century) comes a new outlook. This is the idea of a quantity that, while it isn't conserved, is at least *never wasted*.

The conservation approach had its source in the grand, old metaphysical principle of 'cause equals effect'. This principle was seen (by Leibniz and later by Julius Robert Mayer and Hermann von Helmholtz) as the reason why attempts at perpetual motion were doomed and why a total 'quantity of motion' in the universe had to be conserved. (It is hard to imagine that such a principle could have any detractors and yet d'Alembert referred to it as that 'vague and obscure axiom that the effect is proportional to its cause.'[46])

Then, in 1744, Pierre-Louis Moreau de Maupertuis (we have met him before on his Lapland adventure) introduced another grand metaphysical principle—that nature is economical with her resources and always acts in the most efficient way possible. This meant that certain quantities should be minimized rather than conserved.

It is rare that an idea is really brand new, there are usually antecedents, and this was certainly true for Maupertuis' idea. As early as the sixth century, Olympiodorus had said in his *Catoptrica* that:

Nature does nothing superfluous or any unnecessary work.[47]

Even earlier there had been a maximum principle: according to Greek mythology, Dido had been granted land to found a city on the condition that the land area was limited to the size of a bull's hide. Dido shrewdly cut the hide into very thin strips, joined the strips together, and then made them form the perimeter of a half-circle on the coastline.

In the seventeenth century, Pierre de Fermat (1601?–65) discovered his Principle of Least Time for the path of a light ray. Maupertuis knew of Fermat's Principle and was searching for a mechanical generalization of it. He also took note of Huygens' result that for a mechanical system in equilibrium under gravity, the centre of gravity must be in the lowest possible position.

Out of these considerations, Maupertuis devised a new quantity that should be minimized—the 'action'—and put forward his grand new principle, the Principle of Least Action.[48] For a particle of mass m, speed v, and distance travelled s, the action was defined as mvs. This product had to be a minimum for the path actually taken by the particle, in other words, $\int mv\,ds$ had to be minimized.

We can see now that action has the dimensions of 'energy × time'. We can also see that it is a highly plausible quantity to be the determinant of mechanical activity. Earlier (in Part I), we saw that candidates for the measure of the 'force of a body in motion' were $\int F\,ds$ and $\int F\,dt$. Now, we have a merging of these two measures, as the least action is equivalent to the minimization of $\int F\,ds\,dt$.

Much vaunted by Maupertuis, the Principle of Least Action was the achievement of which he was most proud. He saw it as the universal principle governing almost all natural processes. For example, in his calculus of pleasure and pain, the total happiness was maximized, the total pain minimized, and the measure of each was a product of its intensity and duration. Furthermore, Maupertuis saw his Principle as scientific proof of God's existence (husbanding Nature's resources was a sign of God's perfection). This was audacious, especially in view of the increasingly anti-religious tendencies of the Age of Enlightenment, and some philosophers objected to these theological and teleological implications (how did the particle know which path to follow?). For all this, was the Principle borne out in practice?

Maupertuis was president of the Berlin Academy of Sciences at this time, and had brought together a team of top scientists and the famed mathematician Euler (see Part II) to work in Berlin. Euler checked up on the Principle of Least Action and found that it *did* apply to the motion of a particle moving in a plane and subject only to central forces. He avoided some inconsistencies in Maupertuis' use of the Principle, and noted that the extremum might exceptionally be a maximum rather than a minimum, and also that the total mechanical energy had to be a constant during the motion. In fact, not only did Euler put the whole Principle on a more sound mathematical basis, he had actually discovered it over a year *before* Maupertuis and accorded it overwhelming importance:

> since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of the maximum or minimum does not appear.[49]

However, Euler was reluctant to promote it unequivocally as he could not be certain of its foundations:

> It seems less easy to define, *a priori*, using metaphysical principles, what this [maximum or minimum] property is.[50]

Euler was an unusually modest fellow and he often let others use the results of his labours. However, nothing could surpass his generosity in the matter of the Principle of Least Action. Not only did Euler cede priority to Maupertuis, he actively promoted Maupertuis' case in the notorious König affair. This is an interesting tale, evocative of the spirit and characters of the eighteenth century, so we shall let it divert our attention for a short while.

The König Affair

According to the *Dictionary of Scientific Biography*, Maupertuis was a difficult person: 'spoilt, intransigent . . . Also of unusual physical appearance—very short and always moving, with many tics, careless of his apparel."[51] However, things had been going relatively smoothly at the Berlin Academy of Sciences (in the late 1740s), and we have seen how Maupertuis had invited Euler to join him there. Maupertuis also ensured the election of a former protegé, Samuel König (1712–57), to the Academy. Some ten years earlier, Maupertuis and König, along with Voltaire and Voltaire's mistress, Emilie du Châtelet, had formed a little group that actively promoted Newton's ideas on the continent. They had sometimes met to discuss philosophy at Mme du Châtelet's mansion at Cirey.

When König arrived in Berlin in 1750, he was warmly received. However, relations turned sour when König almost immediately started to attack the validity of the Principle of Least Action and, worse still, to say that it was really Leibniz's idea all along, as evidenced by a letter written from Leibniz to Hermann. Maupertuis was furious and demanded to see the letter. The letter could not be produced—it seems that it went via a certain Henzi of Berne, who had been decapitated.

When Voltaire arrived on the scene (in a distressed state, as Emilie du Châtelet had recently died in childbirth), matters became even worse. Voltaire, who had once flattered Maupertuis with the soubriquet the 'great Earth-flattener' (see Part II), now turned against him and sided with König. King Frederick took the part of Maupertuis, head of his Academy, and generally tried to calm things down but whether due to past jealousies over Emilie or current jealousies over the attentions of the king, relationships deteriorated further. There were also tensions due to Euler's antipathy towards anything Leibnizian. Euler therefore sided with Maupertuis against König's claims of Leibniz's priority.

The affair reached a peak in 1752 with the publication of Voltaire's vicious satire *Histoire du Docteur Akakia et du Natif de St Malo* (known in English as *The Diatribe of Dr Akakia, Citizen of St Malo*).[52] Through the fictional 'Dr Akakia', Voltaire accused Maupertuis of plagiarism, of having a tyrannical rule over the Academy, and other things. The syllable 'kak' would have sounded just as offensive in the eighteenth century as it does today.

Maupertuis became sick and left Berlin for his home town of St Malo, still pursued by vitriolic volleys from Voltaire's pen. Maupertuis never really recovered and died some years later in 1759, in Basle, while stopping off at Johann II Bernoulli's on his way back to Berlin.

The Principle of Virtual Work

Let us now return to mathematics and physics. Euler and Lagrange between them developed the mathematics for minimizing an integral—the 'calculus of variations'. For example, in the case of Least Action, $\int mv\, ds$ had to be minimized. Euler's method was to vary the speed, v, by tiny amounts for each point along the path while Lagrange considered one whole path and then another whole path only slightly different from the first, and so on. In both methods something was varied by an infinitesimal amount and the requirement of minimizing the integral then led to some equations of motion. So much for the mathematics.

As regards the physics, we have seen (in Part III) that Euler needed the injection of Daniel Bernoulli's concept of 'potential live force'. Lagrange, also, wasn't happy that the Principle of Least Action was fundamental enough; he wanted to found mechanics on the most elementary and primitive assumptions possible. For this, he turned to the work of another Bernoulli—to Johann Bernoulli's Principle of Virtual Work (see Part II). This had its roots in the Principle of the Lever which Lagrange took to be—at last—sufficiently fundamental. However, in the final analysis, we shall see that Lagrange's mechanics *was* tantamount to a minimization of action after all.

Joseph-Louis (or Guiseppe Lodovico) Lagrange (1736–1813) was an Italian from Turin. He eventually moved away from Turin, never to return, and worked first in Berlin and then in Paris (upon his arrival he lodged at the Louvre at the invitation of Louis XVI). He worked for the Paris Royal Academy of Sciences until this was dissolved in 1793 by the new revolutionary government. He continued working for the Commission of Weights and Measures and then for the Bureau des Longitudes at the new Institut National (the first meeting was in the Salle des Cariatides in the Louvre), then for the École Normale (it only lasted three months and 11 days) and finally for the École Polytechnique.

Curiously, despite a lifetime's correspondence, collaboration and mutual respect for the other leading mathematician of the age, Leonhard Euler, Lagrange and Euler never met. Euler, as we have seen (Part II), had a hectic family life, permanently surrounded by children and grandchildren. Lagrange, like Euler, was happily married, widowed and then remarried, but expressly desired not to have any children as they would interfere with his work.

Euler and Lagrange did have something in common, though (as well as mathematics)—they both had good powers of self-preservation when it came to the tumultuous political, social, and economic upheavals of the eighteenth century. Euler, on being invited to the Berlin Academy of Sciences, was presented to the queen mother of Prussia. She took great interest in the conversations of illustrious men, but could never draw more than monosyllables from Euler. When he was asked to explain his reluctance to talk, Euler replied 'Madam, it is because I have just come from a country [Russia] where every person who speaks is hanged.'[53] Lagrange had also formulated a prudent rule of conduct: 'one of the first principles of every wise man is to conform strictly to the laws of the country in which he is living, even when they are unreasonable.'[54] Laplace, the other leading mathematician of the age, also shrewdly managed to keep in favour with whoever was in power on either side of the French Revolution. By contrast, some major practical scientists of the age—Lavoisier, Lazare Carnot, and Priestley—had a less detached attitude to the real world, and sometimes suffered the extreme consequences (see Chapter 8). Maybe this points to a difference between the nature of mathematics and science, and/or to a difference between the respective practitioners.

Let's now explain Johann Bernoulli's Principle of Virtual Work. It applies to systems in equilibrium—in other words, where there is no movement between the parts. There has always been a problem with analysing such systems—is the fact that there is no movement an indication that there are no forces present, or are all the forces exactly balanced out? For example—Is the boulder on the mountain stationary because its weight is exactly balanced by an equal and opposite reaction from the ground, or (pretending that we know nothing about gravity) is the boulder merely adopting a position adjacent to the surface of the mountain? What is the difference between a rock-shaped pile of sand and a sandstone rock? Is the tent loosely stacked together (like a house of cards) or are there strong forces in the poles and stays? Is the lever stationary because all forces are balanced (and all moments too) or because there are, in fact, no forces?

Our instinct is to disturb the system slightly and see what happens. We try to roll the boulder, pick up the rock, wobble the tent, and jiggle the lever. In the Principle of Virtual Work, the same procedure is adopted but the disturbance is a 'mathematical jiggling' rather than a real, physical one. A virtual (i.e. mathematically imagined) displacement is applied at all relevant points. If there are any internal forces acting at such points, then 'virtual work' is carried out.

The displacements may be only virtual and in arbitrary directions, but there are nevertheless some conditions: we are only allowed to disturb the system at a given fixed instant in time, to an infinitesimal extent, in a reversible way, and in ways that are consistent with the constraints. For example, we don't go rushing in and break the lever arm or knock it off its perch. Rather, we tweak it, mathematically speaking, and let it rotate infinitesimally away from balance.

Johann Bernoulli outlined the principle in a letter to Pierre Varignon (1654–1722) in 1717 (he called it the principle of 'virtual velocities' rather than 'virtual work'). To a system of variously directed forces in equilibrium, an overall virtual displacement is applied. This results in a set of local virtual displacements at the point of application of each internal force. Each force therefore carries out virtual work that is either positive (the displacement is along the direction of the force) or negative (the displacement is against the direction of the force). Johann put forward his general proposition that the total virtual work is zero:

> In every case of equilibrium of forces, in whatever way they are applied, and in whatever directions they act on [one] another, either mediately or immediately, the sum of the positive energies will be equal to the sum of the negative energies, taken as positive.[55]

It is noteworthy that Johann coined a new term for this virtual work—he called it the *energy*. This is the first time that the word 'energy' is used in physics.

It must be remembered that Johann's proposition is a postulate—there is no way to get in there and experimentally check up on the displacements as the whole thing is virtual. In essence, what is being postulated is that the system responds to a virtual nudge by reactive virtual tremors: the system shifts and resettles itself in just such a way as to counteract the nudge, i.e. all virtual adjustments are taken as acting *in concert*. In this way, the total virtual work comes out to zero. A stronger virtual nudge will elicit a stronger virtual response. This is reminiscent of Newton's Third Law of Motion which is also a postulate. However, it turns out that Newton's Third Law is not as wide-ranging as Newton had hoped—it is generally only applicable to the interactions of rigid bodies and is not as universally applicable as the Principle of Virtual Work.

Let's see how different the variational approach is from the Newtonian strategy for statics. Consider pitching a tent and wanting to know whether it is stable. The Newtonian approach would be to calculate the forces at various key points (the compression in the poles and the tension in the stays as felt at the vertices of the tent, the reaction of the ground at the pegs, and so on). One would then check that the total force was zero at each point. The variational approach, on the other hand, is to mathematically disturb the system very slightly and then demand that the tent continues to stay up.

For the reader who is familiar with the techniques of differential calculus, the Principle of Virtual Work postulates that the equilibrium state is the same, to first order, as all the states that are infinitesimally nearby (with the above conditions of reversibility etc.). The equilibrium state is therefore at a stationary point of the virtual work function. (We assume that the work can be described in functional form.) In order to know whether this stationary point is in fact a maximum, a minimum, or a saddle point, we must investigate the second-order variations. When this is done we usually find that the work function is at a minimum with respect to second-order variations. After noting that 'work done' and 'potential energy' are synonymous terms (see Part II), we arrive at the familiar result that at equilibrium the potential energy is at a minimum. This, finally, can be checked up on experimentally, and provides support for the correctness of the Principle of Virtual Work. What is curious is that while the variations in the work are of a virtual kind, we finally arrive at the very real result that the actual potential energy is at a minimum.

One criticism that can be levelled against the Principle of Virtual Work is that it requires us to know what the constraints are in order to make sure that the virtual displacements are in harmony with them (for example, in the case of the lever, we can displace the lever ends up or down, but we don't allow the lever-arm to bend or get longer or shorter). The system must be examined experimentally first so that it can be understood, and the degrees of freedom (explained later) identified.

But in the Newtonian method, also, prior knowledge of the system is needed. One must be able to identify and quantify all the relevant forces. For a complicated system with many constraints, the Newtonian method is usually too cumbersome or even insoluble without simplifying assumptions. For example, how does one quantify the forces maintaining the rigidity of the lever-arm? There is no mechanics, either variational or Newtonian, that is so general that you just feed in the positions and speeds and masses, turn the mathematical handle, and out pops the answer. Some physical insight and modelling of the system must be carried out for each new scenario.

So far so good, but Lagrange wanted to formulate a mechanics that could be applied to all systems, *dynamical* as well as statical. For this he brought in yet another principle, that of d'Alembert.

D'Alembert's Principle

Now that we are no longer dealing with statics, not all forces are balanced and there is some residual motion. In fact, we have $F = ma$, in accordance with Newton's Second Law. Jean d'Alembert (for biographical details, see below) then employs a strategy that appears too good to be true. He subtracts *ma* from both sides of the equation and obtains $F - ma = 0$. By this ruse, he has ended up with a combination of forces (the applied force, *F*, and the reversed 'inertial force', $-ma$) which sum to zero. (The adjective 'inertial' means 'due to the accelerated motion of a mass'.) The system may then once again be considered to be in a state of equilibrium (albeit dynamic rather than static) and so the appropriate techniques of solving equilibrium problems may be used. Lagrange goes on to apply the Principle of Virtual Work to d'Alembert's Principle. In words, d'Alembert's Principle says that the accelerated motions of the particles, when these are reversed, act such as to counteract the applied forces.

That such a trite rearrangement of terms ($F = ma$ goes to $F - ma = 0$) should lead to anything new is, at first glance, astounding. What is even more astounding is that d'Alembert's Principle turns out to be the cornerstone of the whole of mechanics, and is the bridge that takes mechanics from the classical realm on to Einstein's General Relativity and on to quantum mechanics. These remarkable consequences will be touched upon later. There are many ironic twists along the way. For starters, the discredited 'force of a body in motion' (see Part I) has been reintroduced, this time in the guise of 'inertial motion'. Secondly, d'Alembert reviled the concept of force, inertial or otherwise, considering it obscure and metaphysical[56]—yet his Principle relies on an equilibrium of such forces. Finally, d'Alembert's Principle uses the content of Newton's Second Law, that a force causes a body to accelerate; yet d'Alembert can hardly bare to admit this into his mechanics, which he tries to establish on purely rational grounds. He writes (in his *Traité de Dynamique*, published in 1743):

> Why have we gone back to the principle [that $F = ma$], which the whole world now uses, that the accelerating or retarding force is proportional to the element of velocity? A principle supported on that single vague and obscure axiom that the effect is proportional to its cause. . . . We shall in no way examine whether this principle [$F = ma$] is a necessary truth or not. We only say that the evidence that has so far been produced on this matter is irrelevant. Neither shall we accept it, as some geometers [Daniel Bernoulli] have done, as being of purely contingent truth, which would destroy the exactness of mechanics and reduce it to being no more than an experimental science. [!] We shall be content to remark that, true or false, clear or obscure, it [$F = ma$] is useless to mechanics and that, consequently, it should be abolished.[57]

That d'Alembert should have been the architect of his own Principle is truly amazing.

Before proceeding with the Principle, we shall skim through some biographical details of this 'sinister personality'.[58] We therefore leave the exalted plane of rational mechanics and land instead at the school of hard knocks—Jean Le Rond d'Alembert was a foundling. The baby was left on the steps of the church of St Jean le Rond, in Paris, on 16 November 1717. The police traced his parentage to the famous *salonniere* Claudine Guérin de Tencin and a cavalry officer, the Chevalier Louis-Camus Destouches. Mme de Tencin never acknowledged her son, but d'Alembert's father arranged for him to be fostered by a glazier and his wife, paid for his education, and left him a small annuity. D'Alembert wrote his best mathematical and literary works while living with his foster mother for 48 years. He finally 'weaned'[59] himself (as he called it) and went to live with his mistress, the famous *salonniere* Julie de Lespinasse.

As well as his work in mechanics, d'Alembert was also a literary figure and joint editor of the famous *Encyclopedie* with Denis Diderot. His introduction to the *Encyclopedie* was an important document in the Enlightenment and a manifesto for the *philosophes*. D'Alembert hardly ever left the perimeter of Paris—even the court and science academy of Frederick the Great at Berlin was too 'boring'[60] to detain him for long. He was slight and had a rather high-pitched voice, but this only seemed to enhance his reputation as a mimic, a wit, and a brilliant raconteur. He died in 1783, aged 66 years.

D'Alembert's version of his Principle camouflaged the bald statement of $F - ma = 0$, as he considered cases where there were *multiple* forces of constraint, such as those maintaining the fixed separations in a rigid body. These multiple forces acted cooperatively and produced a single 'reaction' to the applied force. However, d'Alembert is really the last place to look for elucidation of d'Alembert's Principle, as he refers to 'velocities' and 'motions' when he should be talking about accelerations and forces.

The Principle can best be understood by means of an example. We choose the very example that gave d'Alembert the idea of reversed accelerations in the first place. This was the case of the compound pendulum, already treated by Huygens, in whose hands it was also very telling (this was the problem that first led Huygens to notice that the sum of mv^2 remained constant—see Chapter 3).

We may regard the compound pendulum (Fig. 7.3) as made up of many simple pendula compounded together. A simple pendulum is just

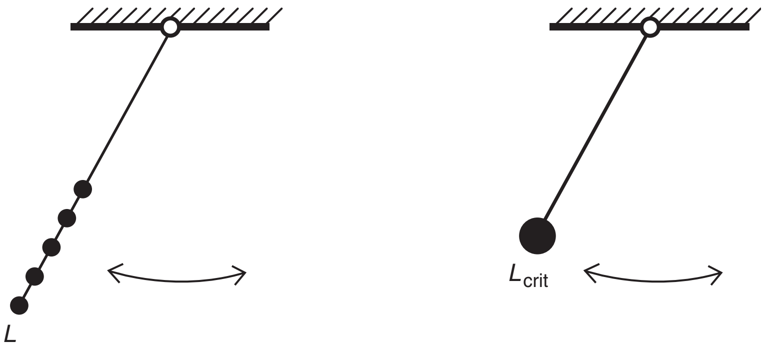

Fig. 7.3 A compound pendulum and an equivalent simple pendulum.

a bob hanging at the end of a cord. Its period (the time of one complete swing) is proportional to the square root of its length (this was well known in the seventeenth century). The compound pendulum swings as if it were just one simple pendulum of some critical length, say L_{crit}, but how big is L_{crit}?

The 'simple pendula' shorter than L_{crit} want to swing with a smaller period than the 'simple pendula' longer than L_{crit}, but all are constrained to move as one compound pendulum (the cord itself is assumed to be weightless and rigid). The point at L_{crit} is neither speeded up or slowed down—it is this critical length that determines the time of swing for the compound pendulum.

In the Newtonian vectorial approach, we examine a given mass-point, m_i, and consider all the various forces acting on it: the force of gravity pulling it straight down, the tension in the cord pulling it up along the line of the cord, and the weight of masses below pulling it down along the line of the cord. We can thus find the net force on the mass-point and calculate its motion by solving $F_{net}(i) = m_i a$. Finally, by considering each mass-point in turn, we can determine the overall motion of the compound pendulum.

Jakob Bernoulli (1654–1705), the first of the Bernoulli mathematicians—the elder brother of Johann and uncle to Daniel—had an ingenious idea. In his seminal paper of 1703, he proclaimed that the compound pendulum could be thought of as a hypothetical lever in which the hypothetical fulcrum coincided with L_{crit}. Masses at lengths shorter than L_{crit} needed to be goaded on, while masses at lengths longer than L_{crit}

needed to be retarded. It was as if small dollops of mass had to be redistributed on the lever to bring it into balance.

The amazing thing is that, as any high school physics student knows, the swing of a pendulum is *independent* of the mass. This shows that this corrective small dollop of mass is truly inertial in origin, that is, it is due to the *motion* of the mass-point (and not to a real extra piece of mass stuck on at various places).

The second amazing thing is that the amount of this corrective dollop of inertial motion is inversely proportional to its distance from the (hypothetical) fulcrum, in other words, it follows the Law of the Lever.[61]

The third amazing thing is that these corrective motions just happen automatically—there is no need to determine any of the internal forces that constrain the mass-points to lie at certain places on the cord. We can simply say (as, in fact, we did) that the cord is 'weightless and rigid' and leave it at that. In other words, the shorthand 'rigid' has all the appropriate constraints implicitly accounted for. This is much simpler than the Newtonian approach in which all the internal forces affecting each and every mass-point must be individually determined.

D'Alembert's great insight was to realize that the same principle operates in *all* mechanical scenarios: the inertial forces (those due merely to the accelerated motions of the particles), when taken in the reversed directions, act as reactive forces, and counteract the applied forces. The usual statement of his Principle, $F - ma = 0$, is not suggestive but it becomes more intuitive when the scenario is more complicated; then, the many and various reactive forces are seen to act in concert, as a *system*. In particular, the internal motions due to any constraint forces are 'lost' or cancel each other out.

All this from someone who reviled the very concept of force—but therein lies the heart of the matter: in considering inertial forces as merely accelerations, d'Alembert freed us up to think of mere accelerations as actual forces. The crucial point is that he treated all forces as equivalent and *ultimately indistinguishable*, whether externally applied, arising from an internal constraint, or from a mass undergoing an acceleration.

For example, it is well known that an accelerating reference frame introduces forces, the so-called 'fictitious' forces. D'Alembert's Principle shows us that these forces are anything but fictitious: in fact, in a fundamental sense, we can't distinguish fictitious forces from the real thing. If we're on a rotating merry-go-round then we are thrown to the outside edge, a ball thrown straight has a curved trajectory, and our stomachs are wrenched so that we feel sick. If there is a thick fog and we can't see the view (in other words, if we don't know that the merry-go-round is rotating), then all the above effects will still occur and we shall have to attribute causes to them other than a rotating frame of reference. D'Alembert's Principle, in treating all forces in an equivalent manner, is showing us that we cannot make a hard and fast distinction between applied and inertial forces: we can't say how much of the force is absolutely given and how much is merely consequential.

Thus d'Alembert's Principle foreshadows Einstein's Principle of Equivalence in his Theory of General Relativity. This Principle asserts that it is impossible to distinguish (locally) between the force of gravity and a uniformly accelerating frame of reference.

D'Alembert's Principle also goes some way to alleviating the metaphysical problems inherent in Newton's Second Law of Motion. Newton had introduced an absolute applied force that resulted in an absolute acceleration (i.e. one relative to an absolutely unaccelerated reference frame). With d'Alembert's Principle we may not be able to remove all these absolutes, but we can at least admit to some arbitrariness in the slicing up between 'force' and 'resulting inertial motion'.

Lagrange's Mechanics

In this tale of many principles, Lagrange recognized in d'Alembert's Principle the golden opportunity to found, at last, the most general system of mechanics. By combining it with Johann Bernoulli's Principle of Virtual Work, he could treat systems *in motion* as well as systems in static equilibrium.

Lagrange's *Analytique Mechanique*[62] (*Analytical Mechanics*) was the pinnacle of eighteenth-century mechanics. William Rowan Hamilton (see Chapter 13) would later refer to it as a 'scientific poem'. It was published in Paris in 1788, just one year after Mozart's opera Don Giovanni and one year before the storming of the Bastille. Certainly there are parallels between the elegance, depth, and beauty of Lagrange's work and that of Mozart, but the links with the French Revolution are more tenuous.

Lagrange's first line of attack had been to use Maupertuis' and Euler's Principle of Least Action (he referred to it as Euler's 'beautiful theorem'[63]). He extended it to apply to many particles and also realized that if ds (the differential of the distance) was replaced by v dt (v is the particle's speed, dt is the differential of time), then $\int mv\, ds$ became $\int mv^2 dt$, or

$\int 2T\,\mathrm{d}t$, where T is the kinetic energy. He and Euler had already developed the mathematics for minimizing such integrals. What would result would be a set of second-order differential equations, one for each particle—the 'Euler–Lagrange' equations.

Lagrange minimized $\int 2T\,\mathrm{d}t$, but then decided not to pursue this approach. The generality of the results that would surely follow from using the Principle of Virtual Work combined with d'Alembert's Principle—both with impeccable foundations (as Lagrange saw it)—was irresistible to him.

Lagrange's style in the *Analytical Mechanics* was most extraordinary: whereas d'Alembert shunned force and energy and just about tolerated mass, Lagrange appeared to shun physics altogether. He proudly announces in the preface that the work contains no figures and no geometrical or mechanical arguments. There are, in fact, no physical explanations of any kind—no mention of particles in collision, rigid rods, elastic membranes, vibrating strings, and so on. Instead, there are 'solely algebraic operations'. The work is as close to pure mathematics and as divorced from physical considerations as a work in mechanics can possibly be.

Let me therefore attempt to put the physics back in. In words, what happens is that, once again, virtual displacements are introduced into a system of forces in equilibrium ($F - ma = 0$). But now the whole thing must be considered as running through *time* as we have a dynamical process rather than the static case. Once again, the total virtual work must come out to zero, and again this is achieved by the virtual work of the 'reactive' inertial forces counteracting the virtual work of the applied forces. However, this balancing out of the virtual work no longer requires the actual motion to be brought to zero—there is some inertial motion left over.

Lagrange introduces some terminology: T and V stand for the kinetic and potential energies, respectively. (Needless to say, he doesn't spell this out—the symbols T and V just sort of arise. As far as I am aware, this is the first time that T and V are given these denotations.) Lagrange also has some conditions: nothing depends explicitly on time, T is dependent on v^2, V must be in the form of a function and must not depend on v, and, finally, $T + V$ (the total energy) must be a constant.

Also, Lagrange introduces his 'generalized coordinates'. Although he doesn't elaborate, this is where all the hard physical thinking comes into each new, mechanical problem. One must examine the scenario carefully (whether it be a compound pendulum, a rotating solid body, in fact, any combination of pulleys, levers, inclined planes, and so on) and then use past experience and general physical nous to determine what are the relevant 'degrees of freedom' of the problem. The degrees of freedom are those features that determine and describe all the possible motions of the mechanical system. The generalized coordinates then map out these degrees of freedom. They should form a complete set (not leave any feature undescribed), but they don't have to be the minimum set possible (some amount of over-determination is allowed).

For example, in the case of the compound pendulum there is only one possible motion (the swing of the pendulum) and therefore only one generalized coordinate is necessary—the angle, θ, that the pendulum makes with the vertical. Note that the generalized coordinate and the corresponding virtual displacement don't have to have the dimensions of length. The generalized coordinates for the problem at hand map out the 'configuration space', a generalization of the usual '3D space' for the usual x,y,z coordinates. Note also that this choice of generalized coordinate has the constraint of the problem built in, so to speak: motions of the individual masses within the pendulum are not allowed and are therefore *not allocated any coordinates*.

With each set of generalized coordinates, the corresponding functions for T and V must be worked out—they will be different for each new problem. These scalar functions, along with specified boundary conditions, will then contain all the physics of the problem. For example, in the case of the compound pendulum (yet again), we must supply various constants of the system, such as: the moment of inertia, I; the total mass, M; the distance, s, between the point of suspension and the centre of mass; and g, the acceleration due to gravity. T is then given by $\frac{1}{2}I(\mathrm{d}\theta/\mathrm{d}t)^2$ and V is given by $Mgs\,(1 - \cos\theta)$.

Once the problem has been set up, we are ready to apply the Principle of Virtual Work to the dynamical case. The virtual displacements are applied and they make two sets of forces, the reversed inertial forces and the applied forces, do virtual work. The requirement is still that the total virtual work must be brought to zero at each instant of time, but now, for a complete solution of the problem, it is necessary to consider a *path in time*. This is so that the dynamic process goes through a whole cycle, or, more colloquially, goes through its paces.

After many pages of algebraic manipulations, Lagrange arrives at the 'Lagrange Equations', the (second-order differential) equations in T and V that describe the motion of the given mechanical system. The amazing thing is that these equations turn out to be the Euler–Lagrange equations *as if* Lagrange had minimized an integral—and not any old integral, but a form of action integral. It seems that Lagrange has unwittingly (?) returned to his earlier approach and to Euler's 'beautiful theorem.'[64] It is hard to credit that Lagrange didn't recognize his own equations popping up (the Euler–Lagrange equations), but he makes no comment (Fig. 7.4).

It turns out that the action integral that has been minimized is $\int(T - V)\mathrm{d}t$ rather than Lagrange's earlier action integral, $\int 2T\,\mathrm{d}t$. They may both be referred to as 'action', as in each case the dimensions are energy × time. The importance to the theme of energy is obvious. The two 'blocks' of energy, kinetic and potential, which we have been at pains to usher in throughout the whole of this chapter, are seen to be at the core of Lagrange's general method of mechanics.

Although it was not until Hamilton in 1834 that mechanics was formulated as starting from this 'least action' principle, we shall take time out from the strict chronological progression of ideas and try to

puisque le signe S est indépendant du signe δ.

Il n'y aura ainsi qu'à chercher la valeur de la quantité $S\Pi m$ en fonction de ξ, ψ, φ, etc. ; ce qui ne demande que la substitution des valeurs de x, y, z, en ξ, ψ, φ, etc., dans les expressions de p, q, etc. (art. 1, sect. II, I^re partie) ; et cette valeur de $S\Pi m$ étant nommée V, on aura immédiatement

$$\delta V = \frac{dV}{d\xi}\delta\xi + \frac{dV}{d\psi}\delta\psi + \frac{dV}{d\varphi}\delta\varphi + \ldots$$

10. De cette manière, la formule générale de la Dynamique (art. 2) sera transformée en celle-ci :

$$\Xi\delta\xi + \Psi\delta\psi + \Phi\delta\varphi + \ldots = o,$$

dans laquelle on aura

$$\Xi = d.\frac{\delta T}{\delta d\xi} - \frac{\delta T}{\delta\xi} + \frac{\delta V}{\delta\xi},$$
$$\Psi = d.\frac{\delta T}{\delta d\psi} - \frac{\delta T}{\delta\psi} + \frac{\delta V}{\delta\psi},$$
$$\Phi = d.\frac{\delta T}{\delta d\varphi} - \frac{\delta T}{\delta\varphi} + \frac{\delta V}{\delta\varphi},$$
$$\ldots\ldots\ldots\ldots\ldots$$

en supposant

$$T = S\left(\frac{dx^2 + dy^2 + dz^2}{2\,dt^2}\right)m, \quad V = S\Pi m,$$

Fig. 7.4 The 'Lagrange equations' and the first appearance of T and V, from *Analytique Mécanique*, 1788 (reproduced courtesy of Kluwer Academic Publishers).

understand the concepts in a physical and qualitative way, ahead of Lagrange's own era. The concepts are difficult, but the consequences of this least action or 'variational' approach turn out to be far-reaching for the future development of physics.

The most remarkable feature of the Lagrange equations of motion (apart from the fact that they work) is that they have exactly the same form irrespective of the problem under investigation: it is always the action integral, $\int(T - V)\mathrm{d}t$, that must be minimized. This is true even though the choice of generalized coordinates varies from problem to problem (it is not even unique for a given problem), and the respective functions for T and V must be worked out afresh each time. To repeat, the functions T and V are different for each new problem and yet the equations of motion remain exactly the same.

The invariance of a set of equations is mysterious and puzzling, but the mystery is partly resolved in the realization that it is a *process*, the 'minimization principle', which is at the heart of this invariance. We are familiar in everyday life with the idea that certain features can remain constant even while other features change. For example, the Khyber Pass is the shortest route through the Hindu Kush from Peshawar in Pakistan to Kabul in Afghanistan, and this is true whether a Mercator's or a Peter's projection is used. Likewise, Mount Everest is the highest mountain, and it doesn't matter whether the scale is in metres or feet. The 'shortest route' or 'highest peak' are superlatives (extremal features) and these are preserved whatever coordinate system, and even whatever sort of map, is used.

But why should it be the 'minimum value of $\int(T - V)\mathrm{d}t$', in particular, that is the relevant invariant feature in mechanics?

In mechanics—say, in a collision between particles—we are not very interested in the actual route and speed of the approaching particles when they are still far enough apart to be considered as 'free'. What is important is their interaction, the event of their collision. We have met this idea before in our discussion of Newton's Laws (Chapter 3). We are interested in the *events*, in what actually happens. As the velocity of a free particle cannot be determined in an absolute sense, then this doesn't count as 'a happening'. In mechanics, the functions T and V determine all the possible details of the system—the possible motions and interactions—like a sort of recipe that describes the whole process in time. However, this recipe still contains some arbitrary features, some features of the motion that could have been different if a different frame of reference or a different map (choice of generalized coordinates) had been chosen. In order to winnow away these map-dependent features and arrive at the events themselves, we must apply a principle, the Principle of Least Action, and this principle will identify the truly invariant features of the mechanical landscape.

Where there are no masses, no energy—we're talking about everyday empty space—it turns out that it is the shortest distance between two points that is the invariant feature between different maps (different 'curved spaces'). When we are talking about mechanics, where we have masses with energy and inertial motion, we shall find that it is the minimum value of the action integral, $\int(T - V)\mathrm{d}t$, that is the relevant invariant feature.

Let us push our intuition even further and ask why should it be exactly $(T - V)$ that is minimized – why not T or $(T + V)$, say? First, let's correct ourselves straightaway and realize that $(T - V)$ isn't very significant: it's the whole mathematical object, $\int(T - V)\mathrm{d}t$, that is the new quantity of importance. 'Acceleration' is a mathematically more complicated and sophisticated object than 'length', but it is describing something more complicated. Now we have an even more complicated object, in keeping with our aim of describing ever more complicated ideas. So, to return to our starting question, why is it $\int(T - V)\mathrm{d}t$ that must be minimized?

Consider minimizing $\int(T - V)\mathrm{d}t$ in the two extreme cases where (a) there is only potential energy and (b) there is only kinetic energy.[65] In case (a), T is zero and therefore we have no motion—we are in static equilibrium. We still have our starting requirement that V is minimized over the whole time interval—but this is exactly what we expect in static scenarios; the potential energy is at a minimum. (We remember that Daniel Bernoulli posited that in static equilibrium a flexed metal band would have a minimum of potential 'energy'.)

Case (b) is even more interesting. Here we have no potential energy and therefore the particle is 'free'. We already know what the resulting motion will be like from Newton's First Law: the particle's speed, v, will be a constant and so will its direction. However, we can get this result in a totally different and novel way, by minimizing T through time.

In this free-particle case, T is given by $\frac{1}{2}mv^2$. As the positions at the start and end of the motion are already determined, we can see straight away that the direction must be constant (any changes in direction would use up more T). It's a bit more subtle, but we can also show that any deviation away from the mean speed would also use up more T. The particle could wait a long time at the starting position, using up no T at all, but then it would have to make an almighty dash just before the time was up, in order to get to the final position on time. We can think of an infinity of other possible variations in the speed, but we are always constrained by the fact that the start and end times are fixed as well as the start and end positions. (Strictly speaking, this applies only in Hamilton's Minimum Principle, which we'll return to in Chapter 13; in Lagrange's Principle, the end-time can vary slightly in order to ensure that the total energy is a constant.). It is ultimately this constraint that determines that the minimal T occurs when v is constant (and equal to the mean speed). (In fact, this follows from the mathematical result that the 'square of the mean' is less than the 'mean of the squares'.)

Now let's consider the more general case where both T and V are present. In the purely static case, the total virtual work is brought to zero. In the more general non-static case, the total virtual work is again brought to zero, but now some resultant inertial motion is allowed. However, while the actual total motion isn't brought to zero, the *excess* of inertial motion is brought as near to zero as possible—not at each instant, but on average, over the whole path in time. This is making the case for the minimization of $\int(T - V)\mathrm{d}t$ at least very plausible.

We are getting close: we must now appreciate that the T-function determines the inertial motions while the V-function relates to the applied forces. So d'Alembert's mixing up of inertial and applied forces now translates into a mixing of T and V, a blurring of the distinction between kinematics (the description of masses in motion) and dynamics (forces generating and altering motion).

We said earlier how the physics of the given mechanical scenario determines the choice of generalized coordinates, so now we shouldn't be surprised if the chosen coordinate system throws physics back at us, so to speak. Take the case of a bead constrained to move on a curved wire (ignoring friction). The bead has only one degree of freedom (motion along the wire) and so only one coordinate is needed—the distance moved along the wire from some fixed point. However, as the bead twists and turns on its way, forces are needed to keep it on the wire. The more massive the bead and the tighter the curvature of the wire, the greater is the force required to keep the bead on the wire as it moves forward. However, these forces aren't explicit, they don't show up in the function V. Rather, these forces are implicit in the choice of coordinates and in the form of the function T (they show up as an extra term in T in addition to the usual $\frac{1}{2}mv^2$). It is this *curvature* in the wire that leads to an extra term in T.

On the other hand, there is also a link between V and curvature. Consider the case of a meteor passing near a planet. It could have its path deflected, or go into orbit or even be drawn into the planet: the stronger the gravitational potential, V, the greater the curving of the meteor's path.

'Curvature' is therefore seen to be both the result and the cause of what actually happens. The T and V components are not just getting mixed up, they are sort of bootstrapping each other into existence. We saw earlier how d'Alembert's Principle anticipated Einstein's Principle of Equivalence. Now again, we find that Lagrange's mechanics anticipates Einstein's Theory of General Relativity, the essence of which is captured in the simplistic slogan 'matter tells space how to curve; curved space tells matter how to move'. Both principles are ultimately tied to Einstein's profound Principle of Relativity—that the laws of physics must rise above the particularities of one or other frame of reference or choice of coordinates; they must describe only the truly invariant physical happenings.

The road has been hard but we have come a long way. There is still one very final step in the argument but it is a conceptually easy one. The question is: why is it exactly $(T - V)$ that must be minimized? The answer is that, as there is this feedback between T and V, we have the danger of each reinforcing the other endlessly, and this could lead to a runaway growth in the total action. This would be highly unphysical – it could lead to perpetual action, which would be even worse than perpetual motion! T and V must therefore act in opposition; they must each act such as to limit the effect of the other. Thus it is $(T - V)$, the *difference* between T and V, that must be minimized through time.

Overview

In this chapter we have seen the evolution of mechanics over 101 years, from Newton's Laws of Motion in 1687 to Lagrange's *Analytical Mechanics* in 1788 via Euler's *Mechanica* in 1736. Leibniz introduced *vis viva*, kinetic energy in all but name, and potential energy slipped in over time. Daniel Bernoulli championed 'energy' before it was fashionable to do so. He developed the kinetic theory of gases and, remarkably, managed to derive an expression for the 'live force' contained within the gas.

There is a tension between Newton's vectorial mechanics and Lagrange's variational mechanics which runs all the way through the eighteenth century and beyond, right up to the present day. In the Newtonian case, we consider an individual particle and watch it as it goes here and there, exposed to different, absolute, externally applied forces. In the variational approach, we look at a whole system, and the scalar functions T and V (the kinetic and potential energies), with appropriate boundary conditions, are the prescriptions that completely determine the progress of the system through time. Which approach is better, which is more fundamental?

We first note that both approaches work. For example, consider a cyclist going round the cambered track at the velodrome. In the Newtonian method we carry out the vector sum of forces and work out the consequent accelerations of the appropriate 'mass-points'. We must take into account the pull of gravity, the reaction of the tyres against the surface of the track, and the 'centripetal force' as the cyclist takes the curve. In the variational approach, we find that in order for the kinetic and potential energies of the cyclist to counteract each other and come out to a minimum value on average through time, the cyclist must climb the track to an intermediary height, somewhere between the highest and lowest extremes of potential and kinetic energy.

It is chiefly with regard to the systems aspects that the two approaches differ. In the Newtonian outlook, the problem has been broken into the participating mass-points and the forces acting on them. There are no precepts such as 'a bicycle'. These Newtonian concepts—masses and forces—feel more elemental and intuitively comprehensible than kinetic and potential energy. And when a handful of mass-points are considered together, isn't this all there is to a system? Not necessarily. In the variational mechanics, the masses don't simply inhabit space and time and wait for a force; rather, the sharp distinction between masses and forces on the one hand and space and time on the other hand has softened. New precepts blending massy and space-time aspects (the functions T and V) are brought in, and these are the ingredients that make up the system.

In this way, the problem of action-at-a-distance is avoided: the system can always be made large enough to encompass any distance required. Moreover, the Newtonian large-scale view is replaced by a local 'field' view and all effects are generated by 'variations' which by definition must be small.

All of this feels unnecessarily complicated in the case of, say, two point-masses attracting each other by a central force—and so it is. Likewise, as children we have need only of integers and real numbers would be an obscuring complication. Initially (in the seventeenth century), Newton had to struggle for his world-view to gain acceptance. Now, on the other hand, we have been led to believe that Newton's force view covers all possibilities. Indeed, Newton's attempt at reducing all physical phenomena to just three laws of motion was probably the most remarkable single leap forward that has ever happened in physics. If we had only to deal with a small number of neutral point-particles then, yes, Newton's vectorial mechanics would answer the case and there would be no need for 'energy'. But we shall find that our physical world—containing moving charges, continuous media, statistical assemblies, microscopic particles, and large gravitating masses—is better served by the systems or energy view.

This is already anticipated in the mechanics of Lagrange. Here, there is a continual interplay between the functions T and V such that within the given system the total energy $(T + V)$ is a constant and $(T - V)$ is minimized through time. But there is a price to pay for the versatility of the new systems view. It is that the new concepts, T and V (the 'blocks' of energy), are complicated, differ from system to system, and have lost the immediate intuitive directness of the familiar forces and masses.

Let's survey the advantages and disadvantages of the systems/'energy' view.

The first advantage is that the generalized coordinates may be chosen in such a way as to incorporate knowledge of any pre-existing symmetries and built-in constraints. This leads to an enormous simplification. For example, we can model the lever arm not just as a specific collection of mass-points but as a 'rigid body' with all that that implies. We then have no need to consider the vector forces that maintain this rigidity.

Secondly, as we have already said, in the systems view effects are always transmitted by *local* variations and yet the system is always sufficiently extensive. Certain philosophical problems are thus avoided.

Third, the systems/energy view is better adapted to comply with the principle of the relativity of motion, as discovered by Galileo and reasserted by Newton (and later by Einstein; see Chapters 17 and 18). In order for a meaningful mechanics to emerge, something absolute about motion has to be extracted. Newtonian mechanics deals with this by looking at the *change* in the motion (the acceleration) of a body as it is subjected to a force. As a change involves a comparison of motions, then this change can be determined absolutely even while the individual motions themselves are relative. But we shall find that the Newtonian stratagem doesn't always work: sometimes accelerations cannot be determined absolutely. However, there is another route to the absolute. In the variational mechanics, a system is defined and then all motion is absolutely determined *within* this system. This approach proved to be a more robust way of satisfying the principle of relativity.

As regards disadvantages, the systems view lacks intuitive simplicity: forces have become 'generalized', and space and time may be melded together or may be 'massy' and chopped into minute components. One of the hardest adjustments to make concerns the evident directionality of motion—how can mere scalars, T and V, hope to capture this directional quality? Upon deeper consideration, we see that direction *is* accounted for in the directions implicit in T and V and in the kinematical constraints. For example, the bead *twists* and *turns* as it moves along the curved wire. In other words, there are directions but all are within the system, none are absolute. (There is also the direction in time, either forwards or backwards, but that is another story—for the time-being.)

Another disadvantage is that T and V must be in functional form in the Lagrangian mechanics. Dissipative effects (such as friction, air resistance, etc.) can more readily be treated in the Newtonian force view, as they can often be lumped together and modelled as one 'dissipative force'. (For example, a marble falling through treacle can be modelled as a body subject to one opposing 'resistive' force.) However, we shall find that when we arrive at the laws of thermodynamics (Chapter 16), then dissipative effects *can* be treated within a systems view and herald another 'block' of energy. This approach is profound and ends up leading to a new cosmic law of nature (the Second Law of Thermodynamics), and is an endorsement of the systems/energy view.

In both the force view and the systems view there are perplexities when extreme cases are considered:

Case (1): there are no forces, only 'kinematics' (no V, only T). For example, in Newton's focus on an individual body it seems paradoxical that the kinetic energy of an isolated body can have its value changed or even brought to zero depending on the reference frame from which it is viewed—where has the kinetic energy gone to? We shall find (in Chapter 18, 'Difficult things') that the 'paradox' is a question of our bias: the body, whilst isolated, is nevertheless crying out to be tied to *us*. Even our use of the phrase 'at rest' is loaded with anthropic connotations. Another example concerns perfectly elastic collisions between hard particles. In the near-infinite number of such collisions in an ideal gas (see Chapters 10 and 16–18), the collisions appear to serve no other function than to 'randomize directions' within the gas.

Case (2): the case of static equilibrium (an equilibrium of forces; pure V, no T). In both the force and the energy view, the system must be disturbed slightly in order to know what's going on. Once again, our difficulty is partly a question of bias: like snakes that detect only movement, we can't help giving more importance to kinetic energy than to potential energy. For example, we consider potential energy as kinetic-energy-in-waiting, but rarely the other way round.

Some problems are endemic to both approaches. For example, both the force view and the energy view share the problem of 'circularity'. Force and mass are introduced together (in Newton's Second Law) and so the definition of one is dependent on the definition of the other. Likewise, in the systems view, we have the problem of defining which effects are within and which are without the system. Resolution in both views comes from the innumerable experimental observations where the consistency of the initial definitions is supported.

This brings me on to teleology. This really is a red herring in both the energy and the force views: the system is no wiser in 'choosing the right path' than is Newton's individual particle in 'knowing how to solve' Newton's equation, $F = ma$, in order to 'know where to move to next'.

To summarize, the Newtonian approach considers the viewpoint of an *individual* (particle or body) as it goes along its path in life, subject to all the 'slings and arrows' (forces) of fortune. This approach is more intuitive but brings in metaphysical problems of absolute force, space, time, and mass. In the energy view, the *systems* view, T and V are not absolute, but their difference, the 'action', *is* an absolute invariant when it is minimized through time. Thus energy and time together are the determinants of 'what happens'.

The blocks of energy, T and V, introduced in this chapter are both types of mechanical energy. Mechanics is rather a contrived discipline. Throughout this chapter we have mostly ignored friction, air resistance, and other 'losses'. There is this sense in which Lagrange's mechanics only bites off what it can chew—it only deals with systems that can be *modelled* (smooth changes in the coordinates etc.). The puzzle is that these concepts, *T* and *V*, end up having a more cosmic role, outside the narrow confines of classical mechanics. Staying within the realm of mechanics a case could be made—a rather poor case—that we could get by with Newton's Laws and without even the concepts of *T* and *V*. It is only when we start to look outside mechanics and discover other 'blocks of energy' that we come to realize that *T* and *V* are forms of energy, and that energy is an indispensable concept. So, for now, we leave the arcane world of algebraic manipulation, the scratching of quill on paper, and look for some other 'blocks' of energy in the belching and clanking of engines and the fizz, bang, and crackle of chemical reactions, gases, electricity, and light.

8

A Tale of Two Countries: the Rise of the Steam Engine and the Caloric Theory of Heat

In France this is a tale of one city, Paris, but in Britain it is a tale of the whole country, and especially of the Scottish university cities of Glasgow and Edinburgh and the new provincial centres in the Midlands, Manchester, and Cornwall.

Our main characters are James Watt and Henry Cavendish in Britain and Lazare Carnot and Antoine Lavoisier in France. As far as Lavoisier was concerned, it was the worst of times (he lost his head to the guillotine), but for French science it was the best of times. Science was not merely a by-product of the French Revolution, it was the *chief cultural expression of it*, ahead of achievements in the arts, music, and literature.[1] The list of French scientists and mathematicians from this period (the Revolution to the Restoration) goes on and on: Lagrange, Laplace, Monge, Condorcet, Jussieu, Lamarck, Cuvier, Saint-Hilaire, Bichat, Lavoisier, Berthollet, Biot, Poisson, Gay-Lussac, Fourier, Coulomb, Coriolis, Navier, Poncelet, Arago, Ampère, Fresnel, Legendre, Galois, Lazare and Sadi Carnot, Magendie, and Cauchy.

The scientists in these two countries could hardly have been more different. The French were salaried professionals and worked in the new French Institute and later in the École Polytechnique. The British were individuals—from rich gentleman-scientists to poorly paid laboratory technicians. They worked from home in their own private laboratories or on site at mills, mines, and in the new 'manufactories'. They formed their own societies to discuss 'matters scientific and philosophical', for example, the Lunar Society in the Midlands (so-called because the members met when there was a full moon and they could travel safely in their carriages after nightfall). Paris was nevertheless the scientific hub and everything French was the touchstone of style. For example, Matthew Boulton, Watt's backer and business partner, welcomed the 'lunatics' to his 'Hotel d'amitié sur Handsworth Heath.'[2]

While the French were in the throes of revolution (political and social), something unplanned and unprecedented was also happening in Britain—industrialization. The pattern of life that had persisted for hundreds of years was changed in just a few decades and forever. It was all to do with engines and the generation of power.

Part I Engines

In Lancashire, for example, virtually every stretch of every river and tributary had a water-wheel on it by the end of the eighteenth century. These were to drive the new spinning and weaving machines of the burgeoning textile industry. But in summer millponds can dry up (yes, even in the North of England) and soon steam engines were being used—not to drive the textile machines themselves but to pump up the water to feed the water-wheels. In the Cornish copper and tin mines steam engines were again used as pumps, but here water was the enemy and the engines were to stop the mines flooding. In 1777, there were just two Watt engines in Cornwall; by 1800, there were 55.[3]

Perpetual motion machines had been shown to be impossible, but these proliferating new machines *could* keep on running provided that they had a source of power. Wind, water, horse, manual, and heat (from burning wood or coal) were all exploited. (In one case, even a Newfoundland dog was used to turn a wheel.) While all these 'fewels' were understood as a source of 'power', the idea of some abstracted essence common to all was still not apparent. Matthew Boulton was a new breed—a venture capitalist and a very advanced thinker. He alone had the understanding that power was soon to be of huge *economic* significance. 'I am selling what the whole world wants: POWER,'[4] he wrote to the Empress Catherine of Russia.

James Watt

The young James Watt (1736–1819) loved tinkering with all the instruments in his father's shipwright's workshop on the banks of the Clyde near Glasgow. He knew what he wanted to be when he grew up—a maker of instruments, particularly mathematical instruments. After an arduous apprenticeship in London, Watt returned to Glasgow for health reasons but couldn't obtain employment as the guilds were too strong. Only one door was open to the young apprentice and that was the post of instrument-maker at the University of Glasgow. So it was, around 1760, that Watt came into contact with Professor Joseph Black (Chapter 6).

Black soon came to appreciate Watt's natural talent for any technical problem and their relationship developed into one of professional co-operation and friendship rather than professor and subordinate. As the student John Robison recalled: 'Dr. Black used to come in [to Watt's rooms] and, standing with his back to us, amuse himself with Birds Quadrant, whistling softly to himself…'[5]

Watt's duties included preparing apparatus for lecture demonstrations. One commission was to bring to working order a benchtop model of a Newcomen engine (the workings of the Newcomen engine are described in Chapter 5). Watt became engrossed in this particular project: he not only got the engine to work but started to wonder at all the factors limiting the functioning of the little engine. Why didn't it work as well as the full-sized engine?

Watt observed that after only a few cycles the model engine slowed down, became very hot, and soon ran out of steam and stopped altogether. He thought deeply about these symptoms and read widely. From his reading of Boerhaave, especially, he realized that the model engine, being small, had a large ratio of surface area to volume, and so the surface of the cylinder would have a disproportionate effect on cooling the steam as compared to the full-size engine (Boerhaave had written about such geometric factors influencing the rate of cooling).

These were impressive conjectures but Watt proceeded to check them out. He determined that a cubic inch of boiling water turned into about a cubic foot of steam, and he also measured how much water was boiled away in a typical run of the engine. Thus he found that the engine used up many times more steam than was required merely to fill the volume of the cylinder. He presumed that this excess was needed just to warm up the cold cylinder at the end of each cycle.

Perhaps making the cylinder out of a different material would lead to less waste of steam. Watt determined the heat capacity of various materials (iron, copper, tin, and wood) and experimented with cylinders made out of wood. These did remain at a more even temperature than the original brass cylinder but were mechanically unsatisfactory (the wood warped and cracked). At this stage, Watt had never heard of Black's concept of specific heat.

Another way to economize would be to avoid cooling the cylinder any more than needed just to cause condensation. Watt therefore investigated the optimum amount and temperature of condensing water required. In so doing, he effectively determined the latent heat of condensation of steam—he still hadn't heard of Black's concept of latent heat.

However, Watt found that when he operated the engine with this minimum of cooling water, then the power of the stroke was severely curtailed. Again, after critical thinking and reading (on Cullen's experiments on the boiling of tepid water in *vacuo*; see Chapter 6), Watt realized what the problem was: warm water left in the cylinder after condensation would boil and generate a back-pressure, vitiating the vacuum and reducing the strength of the power-stroke. Once again, Watt diligently followed up this line of enquiry. He used Papin's 'digester' (Chapter 4) to find the law of the increase of steam pressure with temperature, plotted the results as a graph—a very rare procedure in those days—and extrapolated back to find the pressure corresponding to the temperature of the leftover condensing water.

At this point, Watt faced a dilemma. For economy (the amount of coal burned in the boiler) the cylinder needed to be kept as hot as the steam, while for maximum power it needed to be cooled down once every cycle: but how could the cylinder be kept hot and cold at the same time?

Then, in 1765, in a moment of deep reflection (apparently he was walking across Glasgow Green on a Sunday afternoon),[6] Watt suddenly thought of the solution: he had to divert the steam from the working cylinder into another *separate* chamber, kept permanently cold, where the steam would be condensed. Legend has it that Watt's 'eureka moment' was consequent upon his first hearing of Black's theory of latent heat, but Cardwell demonstrates convincingly that this is nonsense.[7] It seems that Black and Watt simply pursued their own researches at the same place (the University of Glasgow) and at the same time (late 1750s to 1760s) but quite independently of each other, at least during these early years.

The advantages of separate condensation would not only be an enormous saving in fuel (coal for the boiler) but a more powerful engine, and one where the interval between power strokes was much reduced as there was now no need to heat up and cool down one and the same cylinder every cycle. Thus, in 1764, was born Watt's revolutionary 'condensing steam engine'. (We shall not detail the history of its manufacture.)

Watt didn't leave it at that, but continued to improve his engine. He diverted some steam from the boiler so that steam rather than cold atmospheric air would drive down the piston at the end of each cycle; and he continuously throttled down the steam as it escaped into the separate condenser, so that at the end of the stroke every last wisp of steam pressure had been exploited. This became known as Watt's method of 'expansive operation' and in the hands of Sadi Carnot was to become a crucial thought experiment in the founding of thermodynamics (Chapter 12).

In order to put 'expansive operation' into practice, Watt needed to continuously monitor the steam pressure—but everything inside the engine was hidden from view. Watt commissioned his assistant, John Southern, (in 1796) to come up with a solution.

Southern's solution was ingenious. Over a fixed sheet of squared paper, a pen was held in a clamp that moved vertically with the piston and horizontally by a pressure gauge. The resulting 'indicator diagram' showed the variation of pressure with volume.

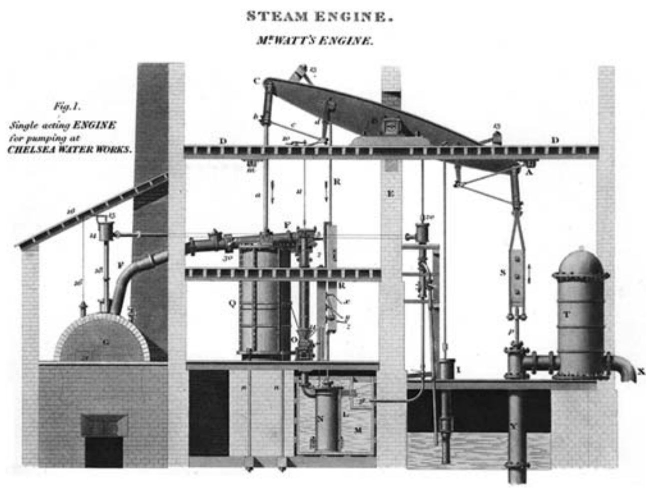

Fig. 8.1 Watt's single-acting steam engine, Chelsea Water Works, *c.* late eighteenth century, from Rees' *Cyclopaedia* (Science Museum, London/SSPL).

What now seems like a simple chart-recording was in fact a revolutionary new practice of inestimable significance. More 'analogue to graphical' than 'analogue to digital', it was a mathematization of the fundamental processes of the engine. Boulton and Watt kept this innovation top secret, but all the competitors were desperate to know how the indicator diagram was generated. The cotton-spinning magnate George Lee complained 'a physi^{n} can give no good account of y^{e} state of his Patient's health without feel^{g} his Pulse, & I have no Indicator',[8] and also 'I am like a man parch'd with thirst in the Expectation of relief, or a Woman dying to hear (or tell) a Secret—to know Southern's Mode of determining Power.'[9]

In later years, Watt further improved his full-sized commercial steam engine by making it 'double-acting' (both up and down movements of the piston constituted a power-stroke) and, in the mid-1780s, able to provide *rotative* power directly (although this design was not fully satisfactory). Both advances required Watt to invent exceedingly clever new mechanisms: the 'parallel motion' and the 'sun and planet gearing'. In his retirement, he told his son that he was 'more proud of the parallel motion than any other mechanical invention I have ever made'.[10]

It is clear that Watt was a genius and not a mere mechanical maestro of the sort that can make anything work just by walking into a room (I don't believe in such sorts anyway)—he always tried to understand the underlying scientific principles. Nevertheless, it's fair to say that Watt's motivation was to improve the efficiency and economy of any design or process rather than study science for its own sake. When working on textile bleaching using chlorine, for example, Watt was horrified that the inventor of the process, the French chemist Claude Berthollet (1748–1822), was 'making his discoveries publick'.[11] Berthollet replied that '*Quand on aime les sciences on a peu besoin de fortune*'[12] ('when one loves science, one has little need of a fortune'). In his memoirs, Watt admitted that he had not had a day free of worries about money until his retirement, aged 80.

(Watt made one significant contribution to pure science—the idea that water was not elemental. Black, by contrast, was of the opinion that water was essentially an 'earth', as it could be reduced to this by evaporation of the liquid parts.)

The revolution brought about by Watt's engine required more than the invention itself but new people (venture capitalists), new social and financial structures, and so on. Watt needed to define quantities such as efficiency ('duty') in order to prove that his engines were better than the competition and also to know how much to charge the customer. He looked back to Parent's treatise (see Chapter 7, Part II) of 1704 on water-wheels—and in doing this he was almost certainly the first to apply the theory of water-engines to a heat-engine. (Watt taught himself French, German, and Italian specifically for the purpose of being able to read foreign texts.)

Parent had defined the efficiency of a water-engine as the percentage of water that could be returned to the starting height. For his steam engine, Watt defined the efficiency (also called the 'duty') as the ratio between the 'work done' and the fuel (bushels of coal) used up. But how should the 'work done' be determined? Watt stuck with common engineering practice and used the specific example of weights lifted through a given height (in fact, pound-weights lifted through a foot), a measure that had been initiated by Descartes over 100 years earlier (see Chapter 3). Finally, Watt introduced a new measure for the *rate* of doing work, the horsepower, which he set at 33,000 foot-pounds per minute. (To commemorate Watt we now have the SI unit of power called the watt, which is defined as one joule of energy per second.) Owners of a Watt steam engine had free installation and servicing, but had to pay a 'royalty' equal to a third of the savings on their previous fuel bills, for example, the cost of hay for horses.

In a dispute about patent infringements, Watt started a lawsuit against the Cornish engineer Jonathan Hornblower (1753–1815), inventor of the compound steam engine. Hornblower asked Davies Gilbert (1811–14), a Cornish MP and mathematician (and later president of the Royal Society), to help defend him. Gilbert (also known as 'Giddy') showed that the area under the graph of pressure versus volume ($\int P\, dV$) was equal to the work done by the engine. Watt, at the trial and becoming alarmed, wrote to Southern: '…they have brought Mr Giddy the high sheriff of Cornwall, an Oxford boy, to prove by *fluxions*, the superiority of their engine, perhaps we shall be obliged to call upon you to come up by Thursday to face his fluxions by common sense…'[13]

To recap, Gilbert showed (in 1792) that $\int P\, dV$ was proportional to the work done by the engine (P is the pressure of the steam or other elastic fluid and dV is the incremental change in the volume). This was an equivalent but completely different measure of 'work done' to the standard contemporary one of 'weight times height'. The formulation $\int P\, dV$ would reappear in the classical thermodynamics of the nineteenth century. This wasn't quite the first time that this expression had been derived; it had been discovered once before, by Daniel Bernoulli almost 60 years earlier, but had passed by unnoticed (see Chapter 7, Part III).

Some Comments

Amontons' and Newcomen's engines were known as 'fire-engines', but by the time of Watt the term 'fire' was replaced by the more prosaic 'steam'. It would still be another 40 years before the steam engine, especially Watt's condensing engine, would be recognized as an engine that ran on *heat* rather than on steam, steam pressure, or anything else.

The heat-engine was a tricky engine to get into production. One could say that the steam engine needed the Industrial Revolution as much as the Industrial Revolution needed the steam engine—in order to provide the coal and the high-precision and high-strength cylinders and pistons, for example. Also, it needed the visionary outlook of Boulton, prepared to risk a fortune with no profits for over a decade, as well as the inventive genius of Watt. However, it still seems strange that the heat-engine, which was so influential in the development of physics, let alone in the industrialization, economics, and politics of the world, should have arisen in just one place—that is to say, in Britain. By contrast, machines using water-, wind-, or muscle-power had cropped up independently all over the globe.

Some have suggested that the steam engine originated in Britain rather than in France because the French had peasants to do the work (before the Revolution) and the aristocrats were only interested in clever devices as 'toys'.[14] Others have argued that England ran out of streams and forests but had a plentiful supply of coal, and so steam power was more suitable than other forms of power.[15] These arguments are persuasive, but from a *physics* perspective, I think that the heat-engine's evolution was unique and serendipitous because it is an inherently difficult thing—one might almost say against nature. In a heat-engine we shall find (in Chapters 16–18) that the randomized motion of the molecules in 'heat energy' must be harnessed and turned into the coordinated motion of a bulk object. Nature has to be cajoled into doing this. This is the substance of the Second Law of Thermodynamics, which emerged from the work of Sadi Carnot (Chapter 12) and was discovered by Clausius and Thomson (Chapter 16). (However, presumably the laws of thermodynamics would still have evolved even if the steam engine had never been discovered...)

It is ironic that these 'unnatural' devices are now the most common sort of engine, the world over. It is even more ironic that the most prevalent type of heat-engine is the reciprocating cylinder engine (in cars) when what is required is direct *rotary* motion. The inefficiency is scandalous and we can only wonder what Watt would have made of it: for example, in my station wagon, running at 100 kilometres per hour, the cylinders come to a dead halt and then reverse direction around 2,000 times per minute.[16]

Lazare Carnot

The engineer-scientist Lazare Carnot (1753–1823) is a rare personage in that he is equally famous as a statesman. He was the 'Organizer of Victory' in the wars of the French Revolution. Later, he served under Napoleon and had to go into exile when the Bourbons came back into power.

In his *Essai sur les machines en general*[17] of 1783, Lazare Carnot was the first to attempt to abstract the idea of a machine or engine to its idealized essentials. He took the conservation of *vis viva* as his guiding physical principle. The chief exponent of this principle had been Daniel Bernoulli, especially with regard to hydrodynamics (see Chapter 7, Part III); and Carnot, although aiming at a general treatment, always used the water-powered engine as his case study.

From the conservation of *vis viva*, Carnot developed his 'Principle of Continuity' for the most efficient mode of operation of a machine: the water had to enter the machine smoothly, continue with no turbulence, and had to leave the machine with as little speed as possible (else this speed could have been exploited to yield yet more power). For example, in the case of an overshot water-wheel, the water should enter the wheel with as little shock to the buckets as possible, should continue with no spray or splashing, and should trickle out at the end.

Lazare sought to analyse machines in terms of the interactions of the 'corpuscles' (infinitesimal elements) of which they were composed. His analysis had shades of Johann Bernoulli's Principle of Virtual Work about it (see Chapter 7, Part II). Lazare considered virtual motions (he called them 'geometric motions'[18]), which were any mathematically possible motions, infinitesimal in extent, and—a novel feature—reversible. For such motions, the Principle of Continuity was satisfied at the microscopic level and the total work done by the machine was the sum of the *vis viva* of its 'corpuscular' parts. Importantly, the concept of reversibility was brought into physics for the first time. Lazare defined it as any 'geometric motion' where the 'contrary motion is always possible'.[19] The idea of reversibility would be crucial in the work of Lazare's son, Sadi Carnot (see Chapter 12).

We have said that the engineers and the mathematicians moved in different worlds—well, Lazare Carnot was the crossover point. For the first time, the *engineers'* concept of work (usually 'force times distance' but sometimes the more specific 'weight times height') began to be incorporated into physics. The process was completed by Gaspard-Gustav Coriolis (1792–1843) who, in 1829, officially defined work as the integral of force times distance. Thus the recognition of 'work', from Amontons and Parent through to Coriolis and via Watt and Lazare Carnot, had taken over 125 years. Was this the first time that an engineering concept had been integrated into the body of physics?

The engineers' 'work' is a block of energy. But even today it sits strangely in physics, relating as it does to such man-made artefacts as machines and having an unusual provenance with no alternative Latin name such as *vis* or *vis viva*. (I think all of us can agree, however, that 'work' is a four-letter word.)

Part II
Theories of Heat

'The French…formulate things, and the English do them',[20] writes Gillispie in *The Edge of Objectivity*. We have already seen an example of this with regard to the science of engines—Lazare Carnot formulated the general theory of machines and Watt built the steam engine. Turning now to chemistry, we find the same trend: the English 'pneumatic chemists', Joseph Priestley (1733–1804) and Henry Cavendish (1731–1810), discovered new gases (oxygen, hydrogen, and others) while the French scientist Antoine Lavoisier incorporated their discoveries into a new system of chemistry and set up the chemical naming conventions that we still use today. We are interested in Lavoisier because of his historic work on heat, *Memoire sur la Chaleur* (*Memoir on Heat*, 1783),[21] which he carried out jointly with the French mathematical physicist Pierre Laplace.

Heat was a topic on which virtually every French philosopher wrote a treatise, from Voltaire and Mme La Marquise du Châtelet to Jean

Hyacinthe de Magellan (grandson of the Portugese explorer) to Jean-Paul Marat, the French Revolutionary who was assassinated in his bath. The picturesque term 'fire' was used more often than the more scientific term 'heat'. The Montgolfier brothers' first hot-air balloon flight occurred in the same year as Lavoisier and Laplace's *Memoir*. Yet, Joseph Montgolfier still felt able to say that their balloon ascended because it was filled with fire, the lightest of the four [Aristotelian] elements.[22]

Antoine Lavoisier

Antoine-Laurent Lavoisier (1743–94) was a Parisian of comfortable middle-class origins. His mother died when he was five years old and he was brought up by an adoring maiden aunt. Following in his father's footsteps, he initially studied law, but a family friend and geologist tempted him into geology and mineralogy, which led into his passion, chemistry.

Lavoisier seems always to have had a great capacity for work. When only 19, he cut himself off from social life, lived on a diet consisting largely of milk, and devoted himself for several months to scientific investigations.[23] His wife tells (in later years) how Lavoisier worked at his beloved chemistry from 6 to 8 a.m. and then again from 7 to 10 p.m., while the rest of the day was taken up with working as a private tax collector (a 'tax farmer') and with his many civic commitments (e.g. supervising the production of gunpowder) and other scientific studies—he was on committees reporting on ballooning, the water supply of Paris, mesmerism (with Benjamin Franklin), invalid chairs, and the respiration of insects. On one day a week, his '*jour de bonheur*',[24] Lavoisier devoted himself entirely to his own researches.

Lavoisier's wife (the daughter of another tax farmer) assisted Lavoisier with his work—translating the works of Priestley and Cavendish into French, recording the results of experiments, and making detailed illustrations.[25] Their apartment and very well-equipped laboratory was at the Arsenal in Paris.

Pierre-Simon Laplace (1749–1827) was one of the major French mathematicians of the Revolutionary period—he claimed that he was *the* major one. (He said this in 1780–81, before Lagrange arrived in Paris in 1787—modesty was not his strong point.) Laplace's chief contributions were in astronomy and probability theory. In physics, Laplace continued Newton's programme of attempting to explain the physical world in terms of attractive and repulsive forces. In around 1807 he founded, with the chemist Claude Berthollet, the Societé d'Arcueil, at their homes on the outskirts of Paris. The physics of the 'Laplacian school' was promoted by Laplace's younger colleagues, Jean-Baptiste Biot and Siméon-Denis Poisson.

We have met Laplace before (Chapter 7, Part II) in connection with potential function theory and the Laplace equation, and we shall meet him again in his corrections to the Newtonian calculation of the speed of sound (Chapter 10).

Laplace was also famous for his founding work in probability theory and for his deterministic views: he speculated that if we could describe the physical world completely at any given time, then the laws of physics would be able to predict the future and determine the past.

Laplace's work in astronomy was written up in his monumental *Traité de Mécanique Céleste*, published in five volumes between 1799 and 1825. Nathaniel Bowditch translated this work into English. He commented that whenever Laplace had written 'It can easily be shown that…',[26] Bowditch knew he had hours of work ahead of him. Laplace's most famous saying arose when Napoleon asked him why God did not appear in the Treatise. Laplace replied, 'I have no need of that hypothesis'—however, the story is apocryphal.[27] Let us now return to the *Memoir on Heat*.

Lavoisier and Laplace's work systematized the science of heat and started off the subject of calorimetry. The work was a collaboration but it was a continuation of Lavoisier's programme of research on heat and it helped him reach his most famous results in chemistry, so we shall continue in his voice alone.

Lavoisier was interested in heat because he wanted to explain combustion and respiration (were they linked?) and vapours and 'airs' (were *they* linked?). Remarkably, even while Watt was carrying out his extensive experiments on water and steam, it had still not been established that all substances can exist in three states—solid, liquid, and vapour. It was Lavoisier who made this explicit, and who saw that it was the quantity of heat within a substance that determined which of these three states was the current one. He saw that vapours and the permanent 'airs' were not fundamentally different—they were all gases. (The *Memoir* was the first time that the word 'gas' gained currency. It came from the Greek *kaos* via van Helmont in the seventeenth century—see also Chapter 5).

In the *Memoir*, Lavoisier starts by describing the two main theories of heat:[28]

Scientists are divided about the nature of heat. A number of them think it is a fluid diffused through nature... Other scientists think that heat is only the result of the imperceptible motions of the constituent particles of matter.

Lavoisier attempted to be impartial and proposed only principles that (he thought) were universal and didn't depend on the nature of heat:

whichever one [hypothesis] we follow, in a simple mixture of bodies the quantity of free heat remains always the same.

But, as with the elastic collisions in Chapter 7, there was some circularity in the definitions (what is a *simple* mixture and what is *free* heat?). Lavoisier continued with his reversibility principle:

If, in any combination or change of state, there is a decrease in free heat, this heat will reappear completely whenever the substances return to their original state; and conversely, if in the combination or in the change of state there is an increase in free heat, this new heat will disappear on the return of the substances to their original state.

This appears incontestable. However, it led to a false assumption—that in the cyclic operations of a heat-engine, all the processes could be reversed. This will be discussed more fully in Chapter 12. ('Free heat' means the heat that can be detected by a thermometer, in other words, the heat that isn't latent or combined.)

How should heat changes be measured? The usual 'method of mixtures' wasn't always appropriate (for example, when trying to determine the amount of heat emitted by a guinea pig over 10½ hours). Laplace devised the ice calorimeter, whereby the subject of the experiment was placed within an ice 'bomb' and the amount of ice that was melted within a given time was proportional to the amount of heat emitted. The design relied on an understanding of latent heat, but it appears that Lavoisier and Laplace knew nothing of Black's work. (They learned of specific and latent heats from Magellan's translation of Adair Crawford's work *Experiments and Observations on Animal Heat*—see Chapter 6.) The ice calorimeter was very successful (in fact, it had already been invented by Black but he graciously acknowledged that Laplace had discovered it independently). Nevertheless, it isn't easy to see how it could have been used in those 'endothermic' reactions where heat is absorbed rather than emitted.

Lavoisier and Laplace carried out three main categories of investigation—the measurement of specific heats, the heats of chemical reactions, and the heat evolved on combustion and respiration.

In the case of the aforementioned guinea pig, the poor creature had to go on a starvation diet, remain in a small basket held at freezing temperature, and have only three inputs of fresh air in 10 hours and 36 minutes. In later experiments (to determine the gases emitted upon exhalation), a further insult to injury occurred when the guinea pig was forced through a trough of mercury on its way to the bell jar. At any rate, the conclusions were important: the exhaled gas was 'fixed air' rather than the phlogistonists' 'vitiated air' and respiration was thus shown to be a form of combustion.

This and all the other ice-calorimeter experiments helped Lavoisier to arrive at his famous conclusion—that all burning, whether by combustion, 'calcination', rusting, or respiration, is *oxidation*. In all cases a component of the air, oxygen, is used up. Hence, Lavoisier had debunked the phlogiston theory (Chapter 4). However, he wrongly thought of oxygen as the 'acidifying principle' ('oxygen' meant 'acid' in Lavoisier's Greek lexicon of chemical terms).

As we have said, Lavoisier claimed to be neutral on the question of the nature of heat but he was, privately, totally committed to the idea of heat as a special fluid. In fact, he needed the heat-fluid as the expansive agent in boiling and evaporation, especially where these processes occurred in a vacuum. (Boiling, evaporation, and even sublimation had been found to occur in the evacuated receiver of the air pump. Where air *was* present, then some liquid could bind with it and become, literally, airborne.)

Lavoisier's views had thus advanced to the point at which he saw heat as instrumental in changes of state and in the elastic properties of gases. He saw a gas as a combination of a gaseous 'base' with the subtle matter of heat. (It could be argued, therefore, that Lavoisier was still a phlogistonist of sorts and had merely transferred the phlogiston from the burning substance to the flammable gas.) Some of Lavoisier's ideas were apparently inspired by an obscure article, '*Expansibilité*', in Diderot's *Encyclopédie*,[29] written anonymously by a government minister, Anne Robert Jacques Turgot (1727–81). (The *Encyclopédie* was a banned book.) What is clear is that Lavoisier's ideas on heat and the gaseous state were completely interdependent.

Only five years later, Lavoisier was unequivocal on the nature of heat. In his famous treatise, *The Elements of Chemistry*,[30] published in 1789, he listed 'caloric' (the subtle heat-fluid) and 'light' as two of his 'primitive elements'. However, taking heat as a subtle fluid did prejudice Lavoisier into making an unwarranted but largely unconscious assumption—that the total heat is always conserved. A fluid substance may be transmitted here and there, sometimes free, sometimes combined, but it is never created or destroyed. But in the operations of a heat-engine we shall find (Chapter 16) that the total heat is *not* conserved; it is instead the even more subtle stuff, *energy*, that is conserved.

When Lavoisier was at the height of his powers, his life was cut short as he fell prey to The Terror and to the guillotine (on 8 May 1794 in the Place de la Révolution, now called the Place de la Concorde). Lavoisier was a committed revolutionary but he came under suspicion due to his work as a 'tax farmer'. Lagrange said, '*Il ne leur a fallu qu'un moment pour faire tomber cette tête, et cent années peut-être ne suffiront pas pour en reproduire une semblable*' ('It took but a moment to make this head fall, but even a hundred years may not be enough to produce a like one').[31]

We shall discuss in a while the reasons why the caloric or subtle-fluid theory of heat was so appealing and became so entrenched, but let us turn first to the tale of a most unusual scientist who swam against the tide.

Henry Cavendish

Henry Cavendish (1731–1810) was born into one of the richest and best-connected families in England: his father was the son of the second Duke of Devonshire, his mother was Lady Anne de Grey, daughter of the Duke of Kent. Henry Cavendish inherited a fortune of over a million pounds but was completely uninterested in wealth. Once, when his banker arrived unexpectedly, Cavendish said 'If it [the money] is any trouble to you, I will take it out of your hands. Do not come here to plague me.'[32]

A career in politics was open to Henry, but he was as uninterested in politics as he was in money. Fortunately, his family gave him free rein to pursue his passion for science. He converted his large residence on Clapham Common (and also the family apartments on Great Marlborough Street) into laboratories, and carried out the experiments on which his fame rests: the discovery of inflammable air (hydrogen), the identification of the dew formed when Priestley ignited a mixture of 'inflammable air' and common air (the dew was water), weighing the Earth, and many others. He was meticulous in his experimenting and achieved unparalleled precision for those times.

Less well known is that Cavendish was also a brilliant theoretician. A hundred years after Cavendish's death, Maxwell was sorting through Cavendish's electrical researches and found that Cavendish had already discovered Ohm's Law, Coulomb's Law, the concepts of electric potential and capacitance, and many other results. In 'pneumatics', Cavendish had also anticipated Dalton's Law of partial pressures and Charles' Law relating the volume and temperature of a gas.

It may have been Cavendish's loathing of publicity and his eccentric personality that led to his keeping so much material unpublished. The recent trend in retrospective diagnosis is to be deplored, but I have to admit that in Cavendish's case he could be the prototype for Asperger Syndrome. There are many anecdotes relating to his extreme shyness; he communicated with his house-servants by notes and is said to have had an extra staircase put in to avoid encountering the maid. Even with scientific colleagues, Cavendish was painfully shy. A fellow of the Royal Society recounts how Cavendish was introduced to a visitor as a celebrated philosopher and forced to listen to a flattering speech: 'Mr. Cavendish answered not a word, but stood with his eyes cast down, quite abashed and confounded. At last, spying an opening in the crowd, he darted through it with all the speed of which he was master, nor did he stop till he reached his carriage, which drove him directly home.'[33]

In 1969, some papers from the Cavendish estate at Chatsworth House came to light for the first time. Here was uncovered Henry Cavendish's 43-page work, 'Heat', probably written in 1787.[34] Cavendish's views were against the contemporary orthodoxy, the subtle-fluid theory of heat, and this may provide another reason why he didn't publish them: 'I think Sir Isaac Newton's opinion, that heat consists in the internal motion of the particles of bodies, much the most probable.'[35]

But what was this internal motion? Even while he was a staunch Newtonian, Cavendish realized that it was not the Newtonian but, rather, the Leibnizian quantification of 'motion'—in other words, $\frac{1}{2}mv^2$—that answered to the needs of heat. He took a body as being made up of an 'inconceivable number'[36] of interacting microscopic particles (actually, too small to be seen in the microscope) and tacitly assumed that the known laws of mechanics would apply on this microscopic scale. He proceeded to an analysis reminiscent of Daniel Bernoulli's—not the latter's kinetic theory, but his interactions of celestial bodies through central attractive forces (a central attraction is one that depends only on relative position) (Chapter 7, Part III). Cavendish considered that the microscopic constituents of bodies would have 'active *vis viva*' (which would affect a thermometer) and 'inactive *vis viva*', which resulted from the relative positions of the particles and was a measure of the 'latent' heat of the body. Both Bernoulli and Cavendish took the conservation of total *vis viva* as the overriding principle.

Thus Cavendish had outlined the first kinetic theory not related to a gas. He was able to use his theory to explain all the phenomenology of heat. First, for the communication of heat between identical bodies at different temperatures, he saw that the motion of particles in the hotter body would be slowed down, while those in the cooler body would be speeded up until the 'particles of both come to vibrate with the same velocity'.[37] For different bodies (e.g. lead and copper), Cavendish was hazy about what was being equalized. He couldn't bring himself to admit to an equality of active *vis viva* (kinetic energy) for particles of very different masses.

Secondly, Cavendish saw that the different particle configurations and strengths of interactions within different materials meant that the proportions of active to latent heat would be specific to each type of material—this explained Black's concept of specific heat. Likewise, a change of configuration during a change of state or chemical reaction explained the latent heats of fusion and vaporization and the 'heats of reactions'.

Finally, Cavendish also explained the heating effect of electricity to his partial satisfaction, but admitted that 'it is an effect which I should not have expected'.[38]

In one class of heat phenomenology, that due to radiant heat, Cavendish readily admitted a material theory of heat. Referring to the experimental investigations of Scheele and of de Saussure (see Chapter 6), Cavendish was convinced by the parallels between heat and light and therefore took radiant heat to be made up of rays of 'heat-particles', by analogy with Newton's rays of 'light corpuscles'. Cavendish also knew of the experiments to determine the momentum of light carried out by the Reverend John Michell (1724–93) and used Michell's results to calculate the work done by light. He found that the light falling on 1½ square feet of thin copper sheet would cause it to rotate and do work at the rate of approximately 2 horsepower.

We can see that in both his motion theory and his radiant theory of heat, Cavendish gave pre-eminence to the principle of the conservation of energy (before energy, as such, had been discovered). Although we need to depart from classical physics in order to quantify the energy of light or of radiant heat, in other respects Cavendish's work tallies pretty well with our present-day views. We now adhere to a motion theory of heat, but in the special case of radiant heat we *do* use the notion of a heat-particle—the photon. However, we now understand that for bodies reaching the same temperature, the final kinetic energy per constituent particle *is* the same, even where the particle masses may be different. Well, not necessarily *exactly* the same. Letting Cavendish have the last word:

> strictly speaking, the values of s & S [active and inactive *vis viva*] . . . must be continually varying, & can never remain exactly the same even for an instant. Yet as the number of vibrating particles even in the smallest body must be inconceivably great, & as the *vis viva* of one must be increasing while another is diminishing, we may safely conclude that neither of them can sensibly alter.[39]

Cavendish had therefore realized that the process of temperature equalization was really a *dynamic* one, operating within an assembly of an enormous number of particles. He was close to, but had not quite reached, the idea of an *average* speed or *average* kinetic energy.

Daniel Bernoulli's work was not picked up by the scientific community and Cavendish's work was hidden from it. The motion or kinetic theory of heat would have to be discovered afresh. In the meantime, we return to the question of why the fluid theory held such a grip.

The Caloric Theory of Heat

The eighteenth century was the heyday of the subtle fluid. Ethers or subtle fluids were invoked to explain gravity, electricity, magnetism, phlogiston, and heat. (In the case of electricity, we could argue that the subtle fluid has never gone away.) Cavendish was an adherent of the phlogiston theory, but instead of thinking of phlogiston as a subtle fluid he identified it directly with his 'inflammable air' (hydrogen). (This at least avoided the problem of having phlogiston as subtle and imponderable as hydrogen has weight and other measurable properties.)

In the case of heat, the idea of a subtle fluid was exceptionally rich and fruitful in explaining almost all the phenomena:

- First, there was Joseph Black's 'equilibrium of heat'—the temperature of bodies becomes equalized in the same way that water (a real ponderous fluid) finds a common level.
- Irvine's theory of heat capacity (Chapter 6) drew on the analogy of a tank holding water to describe a body containing heat-fluid.
- There were analogies made between water-powered engines (waterwheels and column-of-water engines) and heat-engines: in the first case, the water fell between two heights and caused a mill-wheel, say,

to turn; in the second case, the heat-fluid flowed between a hotter and a lower temperature and caused the heat-engine to carry out work.

- From the supposed property that the heat-particles within the heat-fluid are self-repelling (and this had the authority of Newton behind it; see Chapter 5) arose the idea that extra heat would lead to expansion. (Curiously, Cavendish, using his motion theory, didn't find this in the least bit obvious. He could see that the heating of a body would lead to changes in the average separation of the constituent particles, but why shouldn't this lead to contraction as often as expansion?)
- The class of phenomena known later as 'adiabatic' expansions and compressions (see Chapter 10) was suggestively explained using the subtle-fluid theory. (For example, as a gas was compressed the heat-fluid was squeezed out—like squeezing water out of a sponge—and so the temperature rose.)
- In changes of state, extra heat-fluid was required in order to cause expansion (from solid to liquid and from liquid to gas), but also to overcome the cohesive forces binding particles together within the solid or the liquid. We have already seen how Lavoisier required the heat-fluid in order to explain the expansible nature of gases and changes of state.
- For heat being transmitted across a void or from the Sun, the fluid theory had a more ready explanation than the motion theory. The subtle fluid could be transmitted across the void, perhaps as 'rays of caloric', whereas the void by definition contained no constituent particles whether these were in motion or not. We have seen that Cavendish resorted to rays of heat-particles in this instance.

There was really only one category of phenomena that the fluid theory struggled to explain. We'll reveal this in the work of the colourful character, Count Rumford, in the next chapter.

Overview

In this second half of the eighteenth century, we have witnessed the formulation of the abstract theory of machines (by Lazare Carnot), the building of a very real machine (Watt's steam engine), and the emergence of the caloric theory of heat (Laplace and Lavoisier). As usual, the kinetic theory (Cavendish) was developed in parallel but remained hidden from view.

The caloric, or subtle-fluid, theory was mostly taken up by chemists rather than 'physicists'. This led to a subtle error, almost a subconscious one—that the conservation of heat was unquestionably true and was paramount. In chemistry, substances can be variously ground down, boiled away, mixed together, chemically combined, and so on, but always there is the assumption that the starting substances can be recovered if the processes are gone through in reverse. Heat was hypothesized as a material substance, so it was natural to assume that it too would be conserved in all processes. (As we have seen, the conservation of heat was a central tenet of Lavoisier and Laplace's *Memoir*.) We shall find, however, that in the operations of a heat-engine the total heat is *not* a conserved quantity (see Chapter 12).

We have observed that the steam engine was discovered in just one part of the world and for the fundamental *physical* reason that it is hard to tame heat in this way.

Likewise, the motion or kinetic theory of heat was not readily taken up because the ideas were *difficult*, that is, statistical in nature.

'Heat' is statistical, partly because there are a very large number of microscopic particles involved (there are around 10^{24} atoms in an apple, a mug of tea, or a tub-full of air) but also because we are dealing with statistical *processes* (Prévost's 'mobile equilibrium' and the as-yet undiscovered 'mean free path', 'random walk', and so on). By contrast, 'weight' is not statistical: the weight of a litre of hydrogen *is* just the sum of the weights of the individual gas molecules (this was one way in which molecular weights were first determined).

Cavendish and Daniel Bernoulli made the assumption that the laws of mechanics, first discovered on the macroscopic (everyday) scale, would be applicable on the microscopic scale. (Obviously, when these laws were used to describe microscopic interactions, there were no problems due to friction, heat, or other such macroscopic effects.) They appreciated that *average* microscopic velocities were involved, although neither stated this explicitly. The problem was—they simply didn't know how to proceed: what were the speeds when not at the average? How should this be treated mathematically?

We all know how counterintuitive results can come out of statistics—for example, there is the surprisingly high chance of two people at a party sharing a birthday. Some very counterintuitive and unexpected results began to emerge for heat—we shall cover them from Chapter 16 onwards.

9

Rumford, Davy, and Young

Count Rumford

It was Rumford who rumbled the caloric theory of heat. His famous experiment on the boring of a cannon went straight to the Achilles' heel of the caloric theory—the generation of heat from friction. The caloric theory was, remarkably, able to account for virtually all the varied phenomena of heat (see the end of Chapter 8) and was accepted by the Laplacian school and most of the rest of the scientific community. Only in the case of frictional heating were its solutions forced and *ad hoc*. Such a successful and well-entrenched theory was not going to be overthrown by one experiment—and that carried out by an outsider to the scientific community. Rumford was really an outsider to all communities, as a biographical sketch will show.

Count Rumford (1753–1814) was born Benjamin Thompson, in Woburn, Massachusetts, a small rural town near Boston. His family were farmers. His father died when Benjamin was only 18 months old, but his grandfather and uncle left him some land and a small allowance, so he was hardly destitute, despite the script of Cuvier's eulogy: '. . . leaving his grandson almost penniless. Nothing could be more likely than such a destitute condition to induce a premature display of talent . . .'[1]

Benjamin did show talent—an early aptitude in science, self-education, and self-promotion. He read Boerhaave's *The Elements of Chemistry*[2] (see Chapter 5) when only 17 and this gave him a life-long interest in the study of heat. At 19, he married a 33-year-old widow who was wealthy and well connected. At the young age of 20, he became a major in the army on the loyalist (pro-British) side. This resulted in his precipitate departure, first from his home in Concord (formerly, Rumford), New Hampshire, and subsequently from America: 'I thought it absolutely necessary to abscond for a while, & seek a friendly Asylum in some distant part.'[3]

Thus Rumford was one of the world's first asylum seekers and global citizens. From 1776 onwards (i.e. after the American War of Independence) he lived in England, Germany (Bavaria), and France. He had only one return trip to America (to recruit soldiers for his regiment, the King's American Dragoons) and he never saw his first wife again.

While the youth of today are enjoined to consider flexibility as the single most important job qualification, they had better not try to follow Rumford's example. Starting out as a soldier (a major in America, a colonel in England, and a major-general in Munich), Rumford became a spy (for just about every pairing between America, England, Bavaria, and France) and a profiteer. He also was a statesman and diplomat; a nobleman (knighted in England, made a Count of the Holy Roman Empire in Bavaria—he chose to be the Count of Rumford—and made a knight of the order of St Stanislaus in Poland); a philanthropist and social reformer (putting beggars to work, reforming the army, and initiating workhouses, soup kitchens, a 'house for ladies', a home for illegitimates of noble birth, and founder of the 'English Garden' in Munich); educationalist (founding the Royal Institution in London);[4] and designer and inventor (of invisible ink, lamps, candles, a carriage-wheel, a frigate, a life-belt for a horse, the Rumford fireplace, and many other appliances connected with heat—see below).

Rumford was also a ladies' man: he married and separated twice (his second wife was Anne-Marie Lavoisier, Antoine Lavoisier's widow); had a daughter from his American wife and in addition at least two illegitimate children; and had many mistresses (his favourites were Lady Palmerston in England and the sisters Countess Nogarola and Countess Baumgarten in Bavaria).

On top of all this, Rumford was a scientist, chiefly in the areas of heat and light. His interest in science was twofold—practical and philosophical. On the practical side, Rumford was always motivated to improve the 'economy of human life'.[5] He researched the insulating properties of fabrics (and came up with tog ratings) and he invented the coffee percolator, cooking range, roaster, double-boiler, Rumford stove, and, famously, a smokeless fireplace, still the best design after 200 years.[6] He also researched lighting, the 'science of nutrition', and clothing. (He looked somewhat eccentric in the Paris winter, wearing a white hat and coat to better reflect the 'frigorific' rays.)

On the philosophical side, Rumford designed scientific instruments (the photometer, difference thermometer, and standard candle) and contemplated the nature of heat and light.

From early on, Rumford had misgivings about the material theory of heat. The first experiments that alerted him to problems with this theory were the ones he carried out in England, from 1778 onwards, on gunnery and explosives. He made the astute and surprising observation that the gun barrel was hotter when fired without a bullet inside. Moreover, he drew the right but not in the least bit obvious conclusion: when there *was* a bullet, its bulk motion carried away the excess heat. (With the benefit of hindsight, we can see that Rumford's intuition was pregnant with meaning.)

It wasn't until some ten years later, in 1789, that Rumford again returned to his beloved heat studies, and carried out the experiments that have since become legendary in the history of physics. While supervising works at the arsenal in Munich, Rumford noticed that, in the manufacture of cannon, the brass barrel—and especially the metallic chips and scarf—all became very hot when a new cylinder was being bored (it was easier to make cylinders true and of uniform diameter by boring rather than casting: 'rifling' wasn't common practice until the invention of smokeless gunpowder, well into the nineteenth century). Not content to merely observe this effect, Rumford set out to design and conduct his own experiments 'for the express purpose of generating Heat *by friction*'.[7] He arranged for a deliberately blunt borer to be held fixed against a solid cannon waste-head while the latter was turned on its axis by two horses (Fig. 9.1). The steel borer was 'forcibly shoved (by means of a strong screw) against the bottom of the bore of the cylinder'.[8]

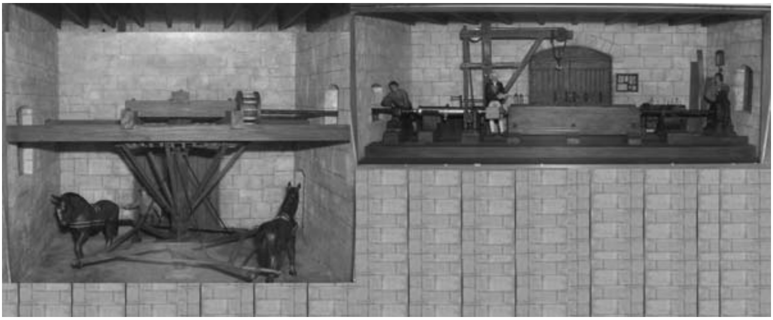

Fig. 9.1 A model of Rumford's cannon-boring experiment (courtesy of the Rumford Historical Association, Woburn, MA).

Rumford found that the temperature of the brass cannon, after 960 revolutions, had gone up by an impressive 70 °F. The caloric theory, in Irvine's version of it (Chapter 6), accounted for this increase by a decrease in the specific heat capacity of the brass chips as compared with the solid brass. The metal chips weighed only 837 grains Troy (approximately 1.9 oz) and Rumford asked 'Is it possible that the very considerable quantity of Heat . . . which actually raised the temperature of above 113 lb of gun-metal at least 70 °F . . . could have been furnished by so inconsiderable a quantity of metallic dust? And this merely in consequence of *a change* of its capacity for Heat?'[9] Furthermore, some earlier experiments had shown that the specific heat capacity of gun-metal was '*not sensibly changed*' by being reduced to the form of metallic chips.

But could the heat have come from the air within the newly drilled cavity? 'I pushed my inquiries a step farther, and endeavoured to find out whether the air did, or did not, contribute anything in the generation of it [the heat]'[10]—by encasing the cannon barrel in a box filled with water. Rumford also took the opportunity of this repeat experiment to continue the tradition of experiment-as-public-demonstration, a prospect he clearly relished. (Galileo had dropped weights from the Leaning Tower of Pisa and Pascal had constructed lengthy barometers of wine and water at the port of Rouen.) In front of an assembled crowd, Rumford waited until (after 2 hours and 30 minutes) the water 'ACTUALLY BOILED! . . . It would be difficult to describe the surprise and astonishment expressed in the countenances of the bystanders, on seeing so large a quantity of cold water heated, and actually made to boil, without any fire.'[11]

No changes in specific heat had occurred, the air did not participate, and neither did the water (the latter had remained chemically unaltered). Another possibility was that heat was being *supplied* to the cannon and borer (say, from the water or via the shaft) at the same time as it was being emitted to the surroundings. But Rumford considered the simultaneous emission and receipt of a material substance as absurd. (For this reason, he utterly rejected Prévost's theory of exchange; see Chapter 6.)

However, the crux of the argument for Rumford was the fact that the heat generated by friction was 'inexhaustible' and so clearly couldn't be a '*material substance*'.[12] In fact, nothing could supply these prodigious quantities of heat 'except it be MOTION'.[13]

One other bit of phenomenology that spoke against the materiality of heat was the fact that heat was weightless, or very nearly so. The prior results on the weight of heat were inconclusive and plagued by the extra weight of water condensing on to cold materials, by convection currents, and by relative changes in buoyancy as bodies cooled.

Rumford cleverly got around these difficulties by comparing the weights of two or more bodies all at the *same* temperature. For example, he compared the weights of water, 'spirit-of-wine' (alcohol), and mercury as these were simultaneously cooled down past 32 °F. This was a telling experiment as the water, upon freezing, lost a large quantity of 'latent' heat as compared to the alcohol and the mercury.

Another telling experiment was the one in which equal weights of water and mercury were cooled from 61 °F to 34 °F. These remained 'in *equilibrio*'[14] despite the fact that the difference in the specific heat capacities of water and mercury meant that the water would have lost around 30 times more heat.

From these experiments, Rumford concluded that heat was indeed weightless, to within the limits of the sensitivity of his balance, 'an excellent balance, belonging to his Most Serene Highness the Elector Palatine Duke of Bavaria'.[15]

In spite of all this, Rumford's cannon-boring experiments cannot be said to have ousted the caloric theory, except in retrospect. Rumford didn't help his cause as he ran a largely negative campaign (i.e. against the caloric theory rather than for the motion theory) and also had some quirky views that were easy to rebut. Chief amongst these was his assertion that fluids are absolute non-conductors of heat. This may have been put forward to bolster Rumford's views of the providential properties of water: 'the law of the condensation of water... [is] palpable proof of the wisdom of the Creator, and of the special care he has taken in the general arrangement of the universe to preserve animal life'.[16]

However, Rumford did further the cause of the motion theory in his choice of lecturers for the Royal Institution in London, which he founded in 1800. Humphry Davy and Thomas Young were two young men, mostly unknown to science, who supported the motion theory of heat and advanced ideas relating to 'energy'. We shall briefly outline their careers.

Rumford's Protégés: Humphry Davy and Thomas Young

Originating from Cornwall, Humphry Davy (1778–1829) is famous for his miners' lamp, the discovery of potassium, sodium, chlorine, and other elements, his inhalation of laughing gas, his associations with the Romantic poets Coleridge and Wordsworth, and finally for *his* choice of protégé at the Royal Institution, the experimenter of genius, Michael Faraday (see Chapter 15).

When only a teenager, Davy had devised an experiment in which he rubbed two blocks of ice against each other until they melted. He satisfied himself that 'It has thus been experimentally demonstrated that caloric or matter of heat does not exist.'[17] He satisfied *himself* but not twentieth-century historians of science who all conclude that Davy couldn't possibly have justified this inference from this experiment[18] (the pressure would have lowered the melting point and the ambient air temperature would also have caused melting). Anyway, Rumford was certainly impressed, and even more so by his young protégé's lecturing style. Davy was a natural performer and pitched the content of his lectures at the right level for his fashionable audience, and also played to the ladies (Fig. 9.2).

Rumford's other new lecturer was the polymathic genius Thomas Young (1773–1829). Young was a child prodigy of the best kind, 'the kind that matures into an adult prodigy'.[19] At age two he was a fluent reader, by six he had read through the Bible twice, and by 13 he was teaching himself Hebrew, Chaldean, Syriac, Samaritan, Arabic, Persian, Turkish, and Ethiopic. He was later a scholar of Greek and Latin, an Egyptologist, and first decipherer of the Rosetta Stone.

Fig. 9.2 'Scientific Researches!—New Discoveries in Pneumaticks!'—Rumford, Davy, and Young at the Royal Institution, by James Gillray (Science Museum, London/SSPL).

In physiology, Young discovered how the eye focuses (by changes in the curvature of the lens rather than the cornea, and not by moving the lens backwards and forwards as in a telescope) and that the eye detects only three primary colours (this had philosophical as well as physiological implications: instead of Locke's evidence of the senses, Young had shown that some so-called raw sense-data, such as the colour purple, isn't raw at all). In engineering, Young is known for the Young's modulus and in physics for the famous 'Young's slits' experiment (which, according to Feynman, brings out the one true mystery of quantum mechanics[20]). Let us now return to 1800 and to the development of the concepts of light, heat, and energy.

Young was the first to explain the curious fact that two sources of light could be combined to yield... darkness. He saw that when two or more light sources were superimposed, they 'interfered', and the resulting patterns of light and dark (for example, as occurred in 'Newton's rings' and 'Young's slits') showed that light was a *wave*. Young thought only in terms of a pressure wave, like sound. However, when the experiments of the French physicist Étienne-Louis Malus (1775–1812), in 1809, showed that light could be 'polarized' (after reflection or transmission through Iceland spar), Young suggested that this could be explained if light was a *transverse* wave, like the standing wave in a violin string.

That light could be a transverse wave was a 'mathematical postulate',[21] but was it physical? Young had his doubts. At this stage (the beginning of the nineteenth century) the idea of waves in empty space was unthinkable, so some sort of medium or ether was implicated. The faster the waves, the more rigid the ether would have to be. The speed of light was known to be very fast, perhaps infinite, and so the 'luminiferous ether' would have to have the contradictory properties of being 'not only highly elastic, but absolutely solid!!!'[22] (these are Young's exclamation marks).

All the above refers to light. However, Young was one of the first to suggest that heat and light were really the same thing. As light was an undulation, then so was heat. This was music to Rumford's ears (almost literally, as Rumford likened the focusing of radiant heat to the tuning of a sound wave to a given note). Young went further and posited that the different colours of light were part of a spectrum in which the frequency of vibration determined the colour. He predicted that heat radiation would occur for vibrations with lower frequencies than the red end of the visible spectrum, and also that invisible high-frequency radiation would be found beyond the violet end of the spectrum. These predictions (in his paper of 1801) occurred almost simultaneously with the experimental findings of William Herschel (1738–1822) in 1800 and Johann Wilhelm Ritter (1776–1810) in 1802, who discovered infrared and ultraviolet radiation, respectively. (Herschel was astonished to see the thermometer rise as it went beyond the red end of the visible spectrum; Ritter witnessed the blackening of silver chloride just after the violet end of the spectrum.)

Connections

All this happened in or close to 1800. The other major achievement in physics at this time was Volta's pile. The Italian physicist Alessandro Volta (1745–1827) stacked alternating discs of zinc and silver, separated by pieces of felt soaked in brine, into a 'pile'. This resulted in the first continuous source of current electricity—a battery. Davy rushed into this new field and was one of the first to show that a chemical reaction (as opposed to the mere contact of dissimilar metals) was essential in producing the current. He also showed that the reverse was true; in other words, that electricity caused chemical reactions to occur, and, ultimately, that chemistry was electrical in origin.

As with heat and chemistry (Chapter 8) and heat and light (above), there were now links between electricity and chemistry. Soon there was found to be a connection between electricity and heat, as in the Seebeck and Peltier effects. (Thomas Johann Seebeck found, in 1821, that when the two ends of a metal bar were held at different temperatures, then a current flowed from the hot end to the cool end. Jean Charles Athanase Peltier found the reverse effect in 1834; in other words, a heat difference was generated when dissimilar metals were connected in an electrical circuit.)

It was long known that electricity and light were linked, as highly charged bodies caused sparking. Also, all three of electricity, light, and heat could be generated by friction. As far back as 1752, Benjamin Franklin (1706–90) had shown that lightning carried electricity (and he lived to tell the tale). Young speculated that the luminiferous and electric ethers were one and the same.[23] However, one of Young's predictions was soon shown to be wrong—that electricity and magnetism would never be linked.[24] In 1820, the Danish physicist Hans Christian Oersted (1777–1851) noticed that a magnetic compass needle flickered when a nearby electrical circuit was switched on; he went on to explore this effect.

Did the connections between heat, light, chemistry, electricity, and magnetism imply that there was some essence that was common to all?

Young lasted only two years at the Royal Institution—his lectures were too abstruse—but in 1807 his masterful two-volume *A Course of Lectures in Natural Philosophy and the Mechanical Arts* was published.[25] Here, he proposed the Greek word 'ευεργεια' (energy) in place of 'living force'. He was the first to use this word in physics since Johann Bernoulli had used it for his concept of virtual work in 1717 (see Chapter 7, Part IV). Although Young introduced the term solely to describe what we would now call *mechanical* energy, Young's coinage was soon applied to diverse 'energies'. Some examples are as follows:

> the relation of electrical energy to chemical affinity is... evident [Davy][26]
>
> different parts of a solar ray, dispersed by a prism, possess very unequal energies... [Biot][27]
>
> Soils... which act with the greatest chemical energy in preserving Manures... [Davy][28]
>
> iodine had less 'energy' than chlorine but more than sulphur [Joseph Louis Gay-Lussac][29]

While the meaning of the term didn't exactly coincide with our modern meaning, the increasing use of the word 'energy' indicated that a new entity was soon to enter the physical stage.

Overview

Rumford's cannon-boring experiment is often cited as the lynchpin in ousting the caloric theory. It *did* do this, but only in retrospect. The caloric theory was well entrenched and it was no good simply showing up a problem with it—an alternative theory had to be put forward. Nevertheless, it is surprising that there weren't more tell-tale cracks within the caloric theory as it was, in fact, false. Who, for example, could doubt the correctness of the motion theory after the invention of friction matches by John Walker in 1827? (He called them 'Congreves' and they were later marketed as 'Lucifers'.)

Young's work showed that light and radiant heat were the same sort of stuff and that the radiation was wave-like rather than being made from rays of 'corpuscles'. The accidental discovery of the thermoelectric effect in 1821 by Seebeck enabled more precise experiments that amply corroborated Young's ideas (especially the works of Jacques-Étienne Bérard, Macedonio Melloni, Edward Forbes, and F. Delaroche[30]). Young had made suggestive propositions, but the true wave theory of light was developed (quite independently of Young) by the brilliant French theoretical physicist Augustin-Jean Fresnel in 1816; and in 1835 André-Marie Ampère developed a wave theory of heat.

But now we have a new dichotomy—instead of a split between the caloric and motion theories of heat, we have a split between the wave theory and everything else: are there two kinds of heat, radiant and ordinary? A possible bridge between the two types of heat lay in the ether. The ether supposedly carried the waves and was itself perhaps made up of tiny particles. However, in 1905, in Einstein's landmark Theory of Special Relativity, the wave was found not to need an ether. Also, after the work of Max Planck, Albert Einstein, and Arthur Compton in the twentieth century, it would be realized that radiation is both a wave *and* a particle (see Chapter 17), and also that there is no ether. All this certainly corroborates the conclusions of Chapter 8—that heat is a difficult thing.

There was increasing evidence, as we have seen, of the links between heat and light, chemistry, electricity, and magnetism—different 'energies'. However, there was still one kind of 'energy' that was not linked to the others. In 1800, the engineer Richard Trevithick (1771–1833) built the first high-pressure steam engine, heralding the age of locomotion and the mass use of heat-engines. Trevithick and his son, Francis, were unable to rouse Rumford's enthusiasm for these engines. Francis comments, somewhat patronizingly, that 'Count Rumford was not quite up to the idea of the new steam engine…still he gave his opinion of a proper fire-place for the boiler of the steam carriage'.[31] Despite promoting the motion theory of heat, Rumford never made the connection between the *microscopic* motion in heat and the *bulk* motion generated by a heat-engine. This is the final link that still had to be made—the link between heat and work.

10

Naked Heat: the Gas Laws and the Specific Heats of Gases

Prologue

The gas laws and specific heats of gases occupy a somewhat dusty corner in physics, but the dust is well and truly blown away when we put the physics into its historical context. We can then appreciate the crucial role of the gas laws in the development of *two* major advances in nineteenth-century science—the theory of energy and the atomic theory. These two theories evolved seemingly independently of each other, except in one arena where they definitely helped each other along: the meeting place was the physics of gases.

Briefly, how the physics of gases helped the atomic theory is as follows.

John Dalton (1766–1844) was the founder of the atomic theory—the theory that atoms are the ultimate constituents of matter and that there are *different* atoms for the different chemical elements (oxygen, hydrogen, gold, sulphur, carbon, etc.). Dalton, in common with most other atomists at this time (around 1800), initially assumed that atomic *size* was the characteristic that determined chemical identity. The effective atomic size, according to Dalton, was given by the size of the envelope of caloric surrounding each hard atomic kernel.

Dalton's espousal of the caloric theory led to his totally static view of a gas: atoms were at fixed positions in a 3D lattice, and were as close to each other as possible consistent with the repulsive force between neighbouring envelopes of caloric (see Fig. 10.1).[1] So, rather than the 'passion-fruit jelly' model of Chapter 5, we now have a new analogy for a gas—polystyrene balls filling a packing case. The polystyrene stacks differently depending on whether the shapes are large or small balls, or even squiggles.

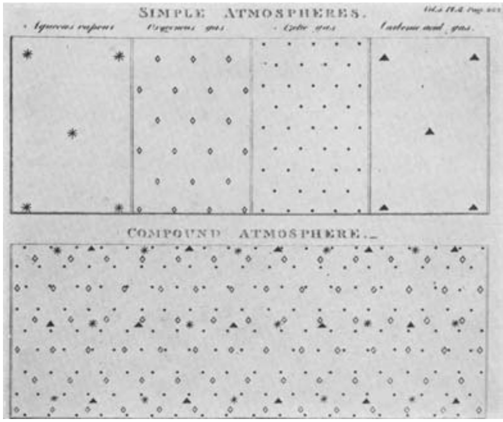

Fig. 10.1 Dalton's 'Simple and compound atmospheres', from 'A new system of chemical philosophy', 1810 (courtesy of the Manchester Literature and Philosophical Society—the 'Lit and Phil').

The first observation, from everyday experience, that puzzled Dalton was the fact that the different components of the atmosphere didn't separate out into layers, with the lightest layer at the top (according to the polystyrene model, each gas, with its own characteristic particle shape and size, should have its own characteristic buoyancy). Other observations in the physics of gases showed Dalton that:

(1) All gases increase in volume to the same extent for a given increase in temperature (when the pressure is held constant).
(2) For a mixture of gases, the pressure of any one gas is completely independent of the presence of the other gases (Dalton's famous law of partial pressures).
(3) The solubility of a given gas in a given liquid depends only on the gas pressure (the higher the pressure, the more soluble the gas).

These *physical* as opposed to chemical results made Dalton re-think his initial ideas. He realized that gas pressure and solubility were purely mechanical effects, and therefore that atomic size and also atomic shape and chemical affinity were all irrelevant. In fact, from the first observation, Dalton surmised that all gases have atoms of exactly the same effective size at a given temperature—in other words, there was the same amount of 'caloric atmosphere' surrounding each atom.

What, then, was the atomic property distinguishing one element from another? Once again, the physics of gases provided the clue. Dalton found that for different gases the solubility did vary according to the gas type: the heavier the gas, the more soluble it was.

Drawing together all these observations from physics and combining them with results from chemistry (chiefly, the relative weights of chemicals combining in compounds), Dalton founded an atomic theory based, for the first time, on *weight*.

Dalton's atomic theory, his ideas on solubility, and his law of partial pressures have all stood the test of time. However, what is astounding and baffling to us moderns is how Dalton's ideas could have survived his totally static and totally wrong conception of a gas—see Fig. 10.1 again and Fig. 10.2.

That his theories came through is because Dalton stumbled across—or, rather, his profound physical intuition led him to—the one salient feature common to both the static and dynamic models of a gas: the temperature alone is decisive. It determines the caloric per atom (static model) or, as we now understand it, the average kinetic energy per gas particle, $<\frac{1}{2}mv^2>$ (modern dynamic model). The gas temperature is

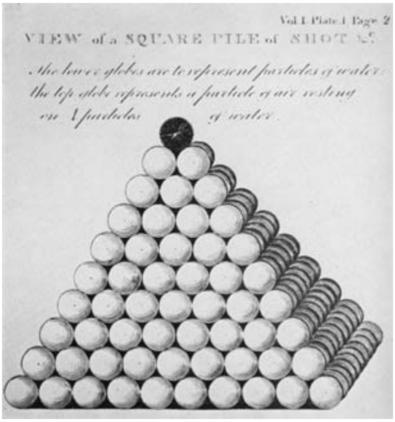

Fig. 10.2 Dalton's 'View of a square pile of shot', from the fourth 'Experimental essay', *Memoirs of the Manchester Literary & Philosophical Society*, 1802 (with the permission of the Manchester 'Lit & Phil'). The caption reads: 'The lower globes are to represent particles of water, the top globe represents a particle of air resting on 4 particles of water.'

oblivious to details such as the particle's, m, mass or speed, v, only the average particle *energy* is important.

One remarkable consequence of this result is Avogadro's Hypothesis, formulated by Amadeo Avogadro (1776–1856) in 1811, one of the most beautiful and simple of all the laws in science. Avogadro realized that the molecule was the basic unit and could be made up of one or more atoms. His hypothesis states that at a given temperature and pressure, a certain volume of gas contains exactly a certain number of molecules. It doesn't matter what the gas is or whether the molecules are like polystyrene balls, squiggles, Maltesers, or Smarties.

Now, if we can just find out *what* that certain number of molecules is, then we'll be able to weigh the gas and find out what the weight of just *one* molecule is. In 1865, some ten years after Avogadro's death, that number was discovered by Josef Loschmidt (1821–95) and named after him. Strictly speaking, it's the number of molecules in one mole of gas; for example, in 16 g of oxygen, at room temperature and atmospheric pressure, there are 6.024×10^{23} molecules.

But wait a moment. If the caloric per atom—in other words, the average kinetic energy—is *the* key to gassy behaviour and is blind to any atomic details, such as the individual atomic weight, how could we ever find out Avogadro's number or any specific atomic features? In other words, what effects could possibly cleave apart the m and the v^2 in the locked casket of $\langle \frac{1}{2}mv^2 \rangle$?

It turns out that there are some effects that do offer a key to this casket. One such effect we have cited already—it is the varying solubility of different gases. A more commonplace effect is the rate of diffusion of gases. Different smells traverse a room at different rates and this depends on individual properties of the smelly molecules. (The transport of smells is, moreover, very hard to explain with a static gas model.)

One final effect I'll mention is the speed of sound in different gases. This speed is related to the average molecular speed, $\langle v \rangle$, in the given gas. For example, in a light gas such as helium the molecular weight, m, is small and so the $\langle v \rangle$ is correspondingly higher. The speed of sound should therefore be higher in helium than in air, and so it is. (This is the reason why people who have breathed in helium talk in squeaky voices.)

That the physics of gases could lead to two such major scientific advances (the atomic theory and the theory of energy) is due ultimately to the fact that gases allow heat-energy to show up in its most naked state (intermolecular forces can usually be ignored in a gas).

Part I
The Gas Laws

Starting with the result, we have

$$PV = nkT \tag{10.1}$$

where P and V are the pressure and volume of the gas, n is the number of particles (molecules of one, two, or more atoms), T is the temperature, and k is a constant (Boltzmann's constant; see Chapters 17 and 18). This law was first discovered empirically (see below) and then derived from the kinetic theory for an idealized gas. A gas is ideal when its particles are point-sized, don't interact, and bounce off the walls elastically. In reality, the gas particles do have a finite size and a small attraction for each other, especially when close together, and so there is a better convergence between theory and experiment when the gas density is low. In the limit as the density goes to zero (in other words, there's no gas at all), then the match is perfect! (See Chapter 18, 'Impossible Things'.)

This law (Equation 10.1) has four variables, so holding any two constant, we have a relationship between the other two. This is summarized in Table 10.1.

We have already described the work of Boyle, Mariotte, and Amontons in the seventeenth century (Chapter 4) and shall now continue with the work of Dalton and Gay-Lussac at the start of the nineteenth century.

John Dalton (1766–1844) was a Quaker and village schoolmaster at Eaglesfield in the Lake District. In 1793, he moved to the smallish town of Manchester to teach at the Manchester Dissenting Academy. He remained in Manchester for 51 years as a teacher, scientist, and meteorologist—he recorded the local rainfall, temperature, barometric pressure, and other weather details every day of his adult life. He never married and his only activities outside science were weekly bowls, an annual trip to the Lake District, and regular church attendance. However, he was affable and had a few close friends (for example, the Reverend Johns) and many scientific contacts in the newly formed Manchester Literary and Philosophical Society, the 'Lit and Phil'.

Table 10.1 The ideal gas laws.

	Relationships	Discovers
n and T fixed	$PV \propto$ constant	Boyle, Mariotte
n and P fixed	$V \propto T$	Amontons, then Dalton and Gay-Lussac
n and V fixed	$P \propto T$	Amontons, then Gay-Lussac
P and T fixed	$V \propto n$	Avogadro's Hypothesis; also implied in Gay-Lussac's work

Manchester was transforming itself as one of the centres of the Industrial Revolution, and Dalton was the first in a line of scientists from Manchester, the 'Manchester school', later to include James Joule (Chapter 14), Ernest Rutherford, and others. Despite this, science in England at this time was still mostly practiced by rich and/or eccentric amateurs. This was in sharp contrast to France where, since the French Revolution, a new breed, the professional scientist, was emerging. This contrast is brought out in the following quote from Sir Henry Roscoe in 1826:

> M. Pelletier, a well-known savant, came to Manchester with the express purpose of visiting the illustrious author of the Atomic Theory. Doubtless, he expected to find the philosopher well known and appreciated by his fellow citizens—probably occupying an official dwelling in a large national building devoted to the prosecution of science, resembling, possibly, his own Collège de France or Sorbonne. There he would expect to find the great chemist lecturing to a large and appreciative audience of advanced students. What was the surprise of the Frenchman to find, on his arrival in Cottonopolis, that the whereabouts of Dalton could only be found after diligent search; and that, when at last he discovered the Manchester philosopher, he found him in a small room of a house in a back street, engaged looking over the shoulders of a small boy who was working his 'ciphering' on a slate. '*Est-ce que j'ai l'honneur de m'addresser à M. Dalton*?' for he could hardly believe his eyes that this was the chemist of European fame, teaching a boy his first four rules. 'Yes' said the matter-of-fact Quaker. 'Wilt thou sit down whilst I put this lad right about his arithmetic?'[2]

As well as the atomic theory, Dalton is famous as the discoverer of (his own) colour-blindness ('Daltonism').

Dalton's most relevant contribution to the gas laws was his fourth 'Experimental Essay',[3] read to the Lit and Phil in 1801. His findings confirmed the investigation of Amontons, 100 years earlier, that all gases expand to the same extent for the same increase in temperature. Dalton was careful to dry the gas and glassware as much as possible using sulphuric acid. He found that the gas increased its volume from 1,000 to 1,376 units between 32 °F and 212 °F. For Dalton, this regular increase in volume was testimony to a regular increase in the volume of the caloric surrounding each separate gas particle as the gas was heated.

At almost exactly the same time as Dalton was carrying out this research, Gay-Lussac was conducting similar investigations in Paris. Gay-Lussac also concluded that the expansion of gases with temperature was uniform, but his experiments were altogether more thorough and professional. This was unsurprising, given that Gay-Lussac was working in a different milieu, an environment that was indeed more professional and where Gay-Lussac was a scientist with a job description and a salary.

Joseph-Louis Gay-Lussac (1778–1850) was, according to his biographer, Maurice Crosland,[4] both a scientist and a bourgeois. He was a protégé of the chemist Berthollet and became a member of the French Institute, First Division, at the young age of 28. Berthollet had recently founded a science society at his house in Arcueil on the outskirts of Paris. He had some rooms converted into laboratories. When the Marquis de Laplace bought the house next door, Berthollet arranged for a door to be put in connecting their two properties, and the new 'Societé d'Arceuil' now had the authority of the great French mathematician at its head. Some other members were Jean-Baptiste Biot, Siméon-Denis Poisson, and the German explorer Alexander von Humboldt. The young Humboldt and Gay-Lussac became friends and carried out some experiments jointly; for example, they measured the combining volumes of oxygen and hydrogen, and going on a balloon ascent to measure the proportion of oxygen at different altitudes.

It was chiefly at Laplace's instigation that Gay-Lussac carried out his early investigations into the physical properties of gases. Laplace, as astronomer, needed to know the factors that affected the refraction of starlight in the atmosphere. Later (see Part II), Laplace needed Gay-Lussac to measure the specific heats of gases. Besides, Gay-Lussac had recently been appointed to the French Institute as a physicist, so he had to find something physical to investigate rather than pursue his beloved chemistry.

Gay-Lussac's research into the expansivity of gases was, as we have said, altogether more professional than Dalton's. He determined that the expansion coefficient was 1/267th per degree Centigrade and that this applied to oxygen, nitrogen, carbon dioxide, air, ammonia, sulphuric ether, and moist air. When Gay-Lussac conscientiously acknowledged some desultory earlier experiments of Jacques Charles', the law of the uniform expansion of gases was mistakenly attributed to Charles ever after (apart from in France).[5]

In any event, Gay-Lussac—as chemist rather than as physicist—achieved lasting fame with a completely different sort of volume relationship. In experiments carried out with Humboldt, he found that hydrogen and oxygen gases combined to form water in a specific ratio by volume: 1.00 volume of oxygen combined with 1.99 volumes of hydrogen. He could have left it at that, but simplicity argued the case for a volume ratio of 1:2. This supposition was given extra confirmation by Gay-Lussac's later experiments from 1807 onwards. He found that whenever gases were the reactants and/or products in a chemical reaction, then the gas volumes were always in a simple ratio. For example, 1 volume of ammonia was made from 1 volume of nitrogen and 3 volumes of hydrogen.

This law of combining volumes of gases, along with Avogadro's Hypothesis, gave wonderful support to Dalton's Atomic Theory. However, Dalton rejected this helping hand, as he couldn't accept the concomitant idea that more than one atom of the same kind could be found in an elementary gas molecule. (For example, for Dalton, ammonia's composition was NH and not NH_3.) So, rather than being grateful for this endorsement of his Theory, Dalton complained of the 'French disease of volumes'.[6] Avogadro's Hypothesis was only accepted 49 years after it was formulated and, unfortunately, four years after Avogadro had died.

Gay-Lussac's researches also helped to bring out the relationship $P \propto T$, which, along with Boyle's Law and Charles' Law (so-called), all confirmed the ideal gas law of Equation 10.1.

We might now ask which are the most important thermodynamic parameters in this equation, $PV = nkT$. Not V or n, as surely no truly fundamental attribute could depend on something so arbitrary as the sample size, whether by volume or by weight. (We shall have to revise this prejudice against 'extensive' quantities later (see the end of Chapter

18, when entropy enters the picture).) That leaves P or T. Black (Chapter 6) had found that bodies all strive to reach an equilibrium temperature, but this applies to pressures too. T demonstrates direction (hot bodies cool down) but the same is true for P (a gas will try and expand rather than contract).

It turns out that T, after all, is the more fundamental thermodynamic parameter. First, every body has a temperature while only gases have pressures. Secondly, and more crucially, two gases at the same pressure are still not in equilibrium until their temperatures have been equalized as well (this is explained in Chapter 18).

Thus temperature wins out over pressure as regards how important and fundamental it is. Only the temperature straddles the macroscopic and microscopic worlds. T is the macroscopic thermodynamic parameter occurring in expressions such as Equation 10.1, but it also relates to the energy of an individual, albeit average, molecule. Pressure, on the other hand, is a meaningless concept when it comes to individual molecules.

Part II
The Specific Heats of Gases

Laplace had discovered a way to improve upon Newton's famously wrong estimation of the speed of sound in air (Newton's result was 10% too low). For this, Laplace needed to know more about the specific heats of gases and he encouraged Gay-Lussac to measure them.

Let's remind ourselves what specific heat capacity is. From Chapter 6, it is the amount of heat needed to cause a one-degree rise in temperature for a given weight of substance. This refers to heat *directly applied*, for example, by a kerosene stove, coal boiler, water-jacket, or whatever. However, in a gas, a temperature rise can also happen in a completely different way—by sudden compression of the gas. This could happen, for example, for air inside a narrow cylinder where a piston is pushed down rapidly (Fig. 10.3). The rise in temperature can be so dramatic that tinder may be ignited (Joseph Mollet[7] demonstrated such 'fire pistons' in the Central School at Lyons in 1804).

So we have the *direct* heating of a gas (say, by flame or 'heat-jacket') and the indirect heating by a sudden change in volume. In the caloric theory, adopted by Dalton, Gay-Lussac, and the 'Laplacian school'

Fig. 10.3 A French demonstration model of a fire piston, early nineteenth century (Science Museum, London/SSPL).

(Laplace, Biot, Poisson, and others), these two processes were thought to be pretty much equivalent, and this led to no end of difficulties, as we shall see.

As in Part I, we shall outline the modern interpretation first. The cases of indirect heating or cooling are called *adiabatic*, meaning that no heat has been transferred between the system and its surroundings. In other words, the system is (presumed to be) enclosed by perfectly insulating walls. The walls of a 'dewar' or thermos flask are a good approximation to such 'adiabatic' walls. (Note that the speed of any volume change is irrelevant in theory but relevant in practice.)

In the case of direct heating, then obviously a transfer of heat across a system boundary has occurred. In one special case this doesn't lead to a change in temperature. This is the case where at the same time as the gas is being directly heated, it is expanding against a piston that does work against the surroundings. We can imagine ideal conditions in which the rate of uptake of external heat exactly compensates for the cooling of the system-gas as it expands. The volume of the gas increases and its pressure falls, but the temperature remains constant. This is called the *isothermal** case. For a given weight of gas at a fixed temperature, all the possible pairs of P and V values join up and define a curve, known as an isothermal, as in Fig. 10.4(a).

We have met this isothermal case before. If we put T = constant into the ideal gas equation (Equation 10.1), then we find that PV = constant (the curves are hyperbolas) and this is also the same thing as Boyle's or Mariotte's Law. Boyle had discovered this law experimentally—and with no direct heating or moving pistons. He had slowly varied the pressure of the gas ('air') by gently raising one arm of a J-shaped glass tube filled with mercury (see Chapter 4) and then noted the accompanying adjustment in the volume of the trapped gas. It is not certain how and to what extent he maintained an accurately constant gas temperature.

We can see how the details of all these various experiments can crucially affect the outcome. But will the two different scenarios—compensated expansion and Boyle's experiment—really map out the *same* curve on the PV graph? The compensated expansion is an ideal set-up, never to be exactly realized in practice (see Chapter 18, 'Impossible Things'). I'm not sure if there has ever been any attempt to check it experimentally and doubtless it would be a tricky business. Nevertheless, for theoretical reasons, we believe that the same curve would indeed be traced out in these two scenarios. The reasons are as follows.

In modern terms, the pressure of a gas arises out of the enormous number of collisions of gas molecules against the walls of the container (around a billion per second per square centimetre for air under standard conditions). Consider a sample of gas at a given temperature and 're-house' it in larger and larger containers. There is no piston, no expansion, and no work is done. As the amount of gas is constant then, as the container size is increased, so the density of the gas molecules and hence the pressure decrease proportionately; and so, from *purely geometric considerations*, we must have PV = constant. So, one of the curves in Fig. 10.4(a) is traced out.

* For future reference, we repeat these definitions. *Isothermal* means at constant temperature, and usually means that the system makes 'thermal contact' with an external heat source through 'diathermal walls'. *Adiabatic* means that the system is contained within insulating or 'adiabatic' walls and there is no transfer of heat between the system and the exterior. In both cases there may still be 'work' done on or by the system by or on the surroundings.

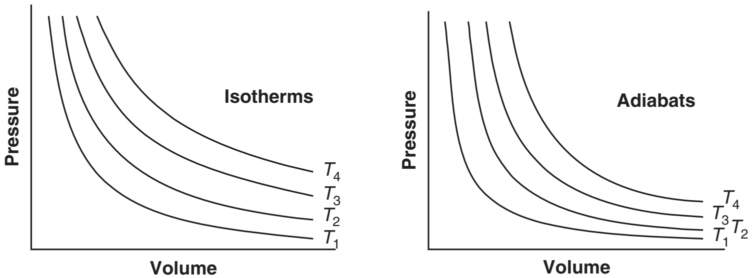

Fig. 10.4 (a) Isotherms and (b) adiabats for an ideal gas.

At another fixed but higher temperature, the gas pressure will be correspondingly higher ($P \propto T$ as in Table 10.1) and so an isotherm higher up will be followed.

If the gas is not simply re-housed but actually expands against a piston, then its volume will increase, its pressure will decrease sharply, and its temperature will fall. However, provided that external heat is supplied such that its temperature is just prevented from falling (maintained constant), then its pressure will be maintained at just such a value that an isotherm (a curve in Fig. 10.4(a)) will again be traced out.

Now for adiabatic changes (the gas pushes back against a piston but the temperature fall is not compensated for by contact with a heat reservoir), the pressure decreases not only for geometric reasons but also because the gas is getting cooler. We therefore have the steeper curves of P against V (see Fig. 10.4(b)). These are called adiabatic curves.

Let us now return to our nineteenth-century narrative.

Gay-Lussac commented that the results of his experiments were not conclusive, as the vessels and thermometers absorbed large amounts of heat relative to that absorbed by the gas itself. Also, the response time of the thermometers was very slow. These considerations were to dog all determinations of the specific heats of gases. Nevertheless, Gay-Lussac tentatively suggested the following trends:

(1) For a given volume of gas, the specific heat for *different* gases is inversely proportional to the weight or density of the gas (e.g. the lightest gas, hydrogen, had the greatest specific heat).

(2) For a given volume of a given gas (in fact, he only showed it for air), the specific heat is less as the pressure is less.

(3) For a given weight of a given gas, the specific heat increases with increasing volume.

The first two suggestions were correct; the last was wrong—but none of them make any sense until we specify the experiments more exactly. In (1), the amount of gas was determined by a given weight at a given volume, whatever the type of gas. Therefore for lighter gases—for example, hydrogen—the sample had a larger number of particles (from Avogadro's Hypothesis). In (2), lower pressures were achieved by releasing some gas. Therefore the sample size became progressively smaller as the pressure was reduced.

In (3), at last the weight of gas was specified—but the third suggestion was wrong. Gay-Lussac kept changing his mind about it, but it was almost forced upon him by the prevailing caloric theory. According to this theory, the subtle heat-fluid, caloric, wrapped itself around each gas molecule like a miniature atmosphere. For a given gas sample, the larger the container volume, the more physical space around each molecule, and the more caloric atmosphere could be accommodated.

The Specific Heat Capacity of the Vacuum

From this wrong supposition came the following extrapolation *ad absurdum*: the specific heat capacity of the vacuum is the largest of all. The eighteenth-century German polymath, Johann Heinrich Lambert (1728–77), believed this and he suggested that suddenly reducing the volume of the void should have a heating effect. Horace-Bénédict de Saussure, Swiss Alpinist and natural philosopher (see Chapter 6), agreed. There were also historical antecedents to this view in the work of Descartes (Chapter 3) and Boerhaave (Chapter 5).

Two French scientists, Nicolas Clément and Charles Desormes, argued the case in a highly original way. They had entered into a competition of the French Institute, First Class, on the specific heats of gases, in 1812. A large spherical vessel was partially evacuated and then left to regain thermal equilibrium. The stopcock was then opened for a fraction of a second and outside air allowed to rush in.

Clément and Desormes claimed that their partially evacuated flask could, in effect, be considered as a mixture between air at atmospheric pressure and a void. The sudden injection of yet more air caused compression and adiabatic heating of the pre-existing components. The relative specific heats of the void and the air could then be found as these two mixed and came to equilibrium (in other words, using the usual 'method of mixtures', see Chapters 5 and 6).

Gay-Lussac was sceptical of all this. He prepared a Torricellian vacuum, tilted the barometer this way and that, and found no changes in temperature as the size of the void was varied. (In the modern view, the vacuum has no assignable specific heat capacity.)

Returning now to the specific heats of gases rather than of voids, another pair of French scientists, F. Delaroche and J.-E. Bérard, carried out a series of investigations that were the last word in accuracy and professionalism at the time. They had entered the same competition as Clément and Desormes (these two pairs were, in fact, the only entrants) and had come first and won the prize of 3,000 francs.

Delaroche and Bérard had gas flowing at a constant rate through a spiral copper tube in a water-bath. In this way, a large volume of gas was used and the uptake of heat was maximized. After calibration, they could work out how much heat had been imparted to how much 'fixed' volume of gas and measure the corresponding temperature increase.

Having thus determined the specific heat capacity of a number of gases, they then extended their measurements to include the heat capacity of just one gas ('air') at just two 'volumes' (in fact, pressures). They found that, taking the specific heat to be one unit at atmospheric pressure, it was 1.2396 units at a pressure of 1 metre of mercury.

From just these two data points, Delaroche and Bérard thought that they had confirmed Gay-Lussac's proposition (3) and with some hubris they wrote:

> Everyone knows that when air is compressed heat is disengaged. This phenomenon has long been explained by the change supposed to take place in its specific heat; but the explanation was founded upon mere supposition, without any direct proof. The experiments which we have carried out seem to us to remove all doubts upon the subject.[8]

This prize error was to reinforce the caloric theory and retard acceptance of the true dynamic theory of heat. Still, it is hard to agree that it was really 'one of the most significant ones [errors] ever made in the history of science'.[9] We shall see, in Chapter 12, that the error was perpetuated in the work of Sadi Carnot, but one has the feeling that Carnot was not totally taken in by it. We shall postpone discussion of this until the appropriate place in Chapter 12.

Apart from the comparison between Boyle's experiment and the case of isothermal expansion, another troublesome duo was the case of a gas expanding due to direct heating (for example, by flame) and a gas heated by adiabatic compression. Was the heat really the same and was the expansion in the first case the inverse of the compression in the second case? Biot thought yes. Gay-Lussac had experimentally found the relationship $V \propto T$ (see Part I) and had measured the expansion coefficient of all gases to be 1/267th for a one-degree rise in temperature. Poisson, using theory, had shown that a (adiabatic) compression of 1/116th caused a temperature rise of one degree. (By this stage, 1800 and after, the French scientists were using the Centigrade scale, now convergent with the Celsius scale.)

Another conundrum was the case of the 'free expansion' of a gas.

Free Expansion of a Gas

Is the case of a gas expanding against a piston the same as the case of a gas expanding into a larger evacuated cavity after the initial containing walls have been removed? (Imagine that the walls have simply melted away.) Today, this is called the 'free expansion' of a gas. The investigation of this effect was a particularly confusing episode, spanning over 50 years. Here are four experiments all giving different outcomes:

(1) First was the experiment of 1807, in which Gay-Lussac's aim was to measure the relative specific heats of gases. He used two 12-litre glass flasks connected by a pipe fitted with a stopcock. The whole apparatus was dried thoroughly with anhydrous calcium chloride and pumped out to as good a vacuum as was possible. The test gas was introduced into one of the flasks and everything was left for 12 hours to ensure stable conditions. The stopcock was then opened and the test gas allowed to rush from the first flask into the empty flask. Air, hydrogen, carbon dioxide, and oxygen were tested in turn, all at roughly atmospheric pressure, and also air at two other lower pressures. The stopcock was specially designed with a variable orifice so that the gases would flow through the pipe at roughly equal rates (otherwise the lightest gas, hydrogen, would flow the fastest). Each flask had its own thermometer.

What Gay-Lussac found was that the temperature of the first flask always fell and the temperature of the second flask always rose by an almost equal amount. (For example, for air at atmospheric pressure, the temperature in the first flask fell by 0.61 °C and the temperature in the second flask rose by 0.58 °C.) The biggest temperature rises and falls were noticed with hydrogen, then air, oxygen, and finally carbon dioxide. Gay-Lussac tentatively concluded that the magnitude of these temperature changes, and hence the specific heat capacities themselves, were in inverse proportion to the gas-weights (for example, hydrogen, the lightest gas, would have the largest specific heat).

(2) We have already described the competition entry of Clément and Desormes, carried out in 1811/12 (published in 1819). Air from outside was allowed to expand into a partially evacuated vessel. The temperature in the vessel rose slightly.

(3) In 1844 James Joule repeated Gay-Lussac's experiment, but with copper vessels instead of glass flasks. There is no evidence that he knew of Gay-Lussac's work. He found no temperature changes upon expansion of the gas from one vessel into the next.

(4) In 1852–53, Joule and Kelvin together carried out a similar experiment, but this time with the gas expanding through a cotton-wool porous plug. The temperature fell. Later (1861), Joule found that the cooling was less marked the higher the starting temperature. Later again, in the twentieth century, it was found that above a certain critical temperature, the inversion temperature, the temperature *rises* as the gas expands.

Leaving you, for the moment, to devise your own answers to these divergent outcomes (all will be explained at the end of the chapter), let us return to the vexed question of specific heats.

C_P and C_V

There was yet one more troublesome duo in specific heat studies. We have talked of the compensated expansion of a gas as it tracks an isothermal curve, but there is another kind of compensation possible—that required to keep the *pressure* rather than the temperature of the gas constant as the gas expands. (We could call this isobaric expansion.) Thus two kinds of specific heat can be defined: C_V, the specific heat at constant volume, is the amount of heat required to raise the temperature of a given weight of gas at constant volume; C_P, the specific heat at constant pressure, is the amount of heat required to raise the temperature of a given weight of gas at constant pressure.

Adair Crawford in 1788 (Chapter 6) was attempting to explain animal heat and was the first scientist to make any measurements whatsoever of the specific heat of a gas. He measured the temperature rise after direct heating, but contained the gas first in an extensible bladder and then, later, in a brass vessel of fixed dimensions. He therefore inadvertently obtained rough measures of C_P and C_V, but of course he had no idea that there were two types of specific heat.

The importance of C_P is that it is explained in quite different ways in the caloric and modern theories. In the caloric theory, the excess of C_P over C_V is due to the 'latent heat of expansion' of the gas: as the gas expands, there is more space for the caloric, the capacity of the gas (to 'hide heat') goes up, and the amount of 'sensible' heat (heat detectable by a thermometer) goes down. In the modern theory, the heat capacity of the gas is *not* dependent on its volume. Rather, the expanding gas does mechanical work in pushing the piston and this work is at the expense of the heat energy in the gas. This recognition—of the actual *conversion of heat into work*—heralded the discovery of 'energy' and will be covered in Chapter 14.

The remarkable thing is that while the qualitative explanations are radically different, the quantitative agreement between the two theories is total. In both theories, $C_P = C_V +$ 'heat of expansion'. Whether this 'heat of expansion' goes to increase a hidden reserve within the gas or whether it is lost as it is converted into work makes no difference whatsoever to the thermometer readings. Moreover, this 'heat of expansion' serves exactly the same role in both the caloric and modern theories: $(C_P - C_V)$ is a constant in both theories; C_P/C_V is another constant, called γ; the adiabatic curves in Fig. 10.2 are given by PV^γ = constant in both theories; and finally, the correction to Newton's calculation of the speed of sound is $\sqrt{\gamma}$ in both theories.

This last was famously worked out by Laplace. In 1816, he published a brief note in which he had closed the gap between Newton's value and the experimental value, but he didn't give his reasoning. In 1823, in the last volume of his magnum opus, the *Mécanique Céleste*, he explained all. (It would be ungenerous to say that in his first paper he looked around for a correction that would fit and then in the intervening years he worked out how to justify it.)

This work was the reason why Laplace had asked Gay-Lussac to investigate the specific heats of gases in the first place. The value of γ was measured by Gay-Lussac and Welter in 1822, using a modification of

Clément and Desormes' apparatus (experiment (2) above) and ever after referred to as Clément and Desormes' Method. Their value of 1.37 brought Laplace's calculations into better agreement with experiment—(so the latter's work could be considered as a re*sound*ing success!).

Dulong and Petit's Law

Before we leave the subject of specific heats, we must look at the influential work of Pierre Louis Dulong and Alexis Thérèse Petit with respect to solids. In 1819 Dulong and Petit measured the specific heat capacities of a number of solids, mostly metals ('*bismuth, plomb, or, platine, etain, argent, zinc, tellure, cuivre, nickel, fer, cobalt, soufre*'). They also knew the relative atomic weights of these solids (from the work of chemists such as Jöns Jacob Berzelius). They then stumbled upon a curious correlation—the product of the specific heat and the relative atomic weight was roughly constant—and realized this meant that the specific heat *per atom* was a constant. They elevated this result to the status of a law:[10]

The atoms of all simple bodies have exactly the same capacity for heat.

This law, while not exact, gave wonderful support to the atomic theory—but Dalton, as usual, was not to be helped.

Dulong further wondered whether there might be a similar law in the case of gases. (The year after the original law was promulgated, Petit died of tuberculosis, aged only 29.) Dulong therefore carried out a comprehensive review of the specific heats of gases, in 1829, and ended up extending Dulong and Petit's Law to the case of gases. He further concluded that as the specific heat per gas molecule was a constant, whatever the gas, then the specific heat for a given *volume* of any gas was also a constant. This is exactly what Dalton and Gay-Lussac had suspected all along (and it also ties in with Avogadro's Hypothesis).

The specific heat capacity, like T (see the end of the prologue), is determined solely by the average kinetic energy per atom and is unrelated to the mass (i.e. the type) of the atom. This explains Dulong and Petit's Law. The same is true when the 'atom' is in fact a gas molecule, and this then explains Dalton's and Gay-Lussac's findings. However, when the gas molecule is polyatomic (for example, carbon dioxide, water vapour, or methane), then there are more avenues through which heat can be taken up apart from purely kinetic. The specific heat is therefore larger in such cases (and this is why the so-called 'greenhouse gases' contribute to global warming).

So it seems that the specific heat of gases, like the temperature, straddles the macroscopic and microscopic worlds. On the one hand, the specific heat is a macroscopic thermodynamic parameter, telling us about a bulk property of the gas—how much heat it absorbs per degree. On the other hand, it tells us about certain details of the actual molecules, such as whether they are made from one, two, or more atoms.

There is yet one more way in which specific heat is interesting. In the early twentieth century, the specific heat (for solids in particular) was one of the first pointers to the new quantum physics. In the classical theory, the specific heat remains constant with temperature. What was observed experimentally, however, was that the specific heat dropped at very low temperatures—it was as if certain avenues for absorbing energy, certain degrees of freedom, were being 'frozen out'. This was explained by Albert Einstein for the case of diamond: the carbon atoms can't just absorb energy any old how, they can only absorb it in discrete chunks or quanta. At extremely low temperatures there isn't enough heat energy available for these quantum energy levels to be populated and the heat capacity is therefore reduced, sometimes discontinuously.

Before we finish this section on the specific heats of gases, we mustn't forget to tie up the loose ends of experiments (1)–(4) above. In modern terms, we explain the findings as follows.

Free Expansions, Again

First, in Gay-Lussac's experiments, the variations in temperature that he observed were entirely due to the fact that he didn't wait long enough after opening the stopcock for equilibrium conditions to be achieved. The effect was more marked for the lighter gases, as these have a higher average speed (see the end of the prologue) and therefore travelled to the second flask more quickly (despite the specially designed valve).

Joule, on the other hand, measured the temperature of the water bath surrounding *both* vessels and, furthermore, he waited until all temperature fluctuations had ceased.

In Clément and Desormes' experiment, as with Gay-Lussac's, the conditions hadn't settled down to a steady state. Also, as there was some air already in the vessel, then the expansion wasn't even approximately 'free'.

The temperature change following the Joule–Kelvin expansion is tricky to explain (and without talking about enthalpy) but is entirely due to the fact that no gas is truly ideal (there *are* interactions between the molecules). First, the gas 'does work on itself', pushing itself through the narrow 'throttle'. The molecules repel each other as the gas is squashed up and the temperature may rise. Then, as the gas emerges into the larger volume, its molecules are further apart and there is some cooling (the molecules attract each other at these larger separations).

Theoretically, in a true free expansion of a gas (no intermolecular forces and no work done pushing back pistons or other gases) the temperature remains constant whatever the volume or pressure of the gas. This counterintuitive result cannot be explained in the caloric theory. In modern classical thermodynamics, we say that the internal energy of an ideal gas depends only on the temperature—it doesn't matter where you are on a given isotherm. Once again, the temperature is appearing as *the* defining thermodynamic parameter.

There is another puzzle—why is there a direction involved? Why do ideal gases tend towards lower pressures and larger volumes? This is mysterious and can only be answered by reference to a new abstract quantity—entropy. We shall meet entropy at the end of Chapter 16 and discuss this in Chapter 18, 'Difficult Things'.

Real Experiments—Some Anecdotes

The experiments all point to the crucial influence of the experimental conditions. To give you a flavour of the ingenuity of these experiments from the early nineteenth century, consider this rather appealing example—Gay-Lussac's method of determining vapour pressure. Gay-Lussac blew a hollow glass teardrop (such as may be seen in the so-called 'Galileo thermometer'), weighed it, filled the bulb completely with the relevant volatile liquid, sealed the neck with a flame, and then reweighed it. This full teardrop was then introduced into a Torricellian barometer so that it rose up through the mercury and bobbed into the vacuum. Heat was then applied from outside the barometer (perhaps with a burning glass?) until the teardrop exploded and the vapours were released.[11]

Taking account of the volume of floating glass shards and of the thermal expansion of the mercury and the vessel, the change in the mercury level was then a measure of the pressure of the alcoholic vapours.

Many of these early experiments were fraught with danger. In 1804, Gay-Lussac ascended higher in a balloon than ever before, 7,016 metres, a record that was not matched for half a century afterwards. He complained of nothing more than a headache. In 1808, he was temporarily blinded by an explosion with the newly discovered potassium. He recovered, but his eyes were permanently affected and he subsequently always wore protective goggles. In 1811, Dulong lost a finger and the sight of one eye during his discovery of nitrogen trichloride. He resumed his investigations only four months later. Needless to say, all chemists ran the risk of breathing in noxious fumes (for example, Carl Wilhelm Scheele checked the smell and taste of his new discovery, hydrogen cyanide).

Overview

One thing that we have learned is that the minutiae of the experiments were crucial, not just with regard to precision but with regard to what was actually being measured. How did Boyle maintain a constant temperature (when checking PV = constant) and Gay-Lussac a constant pressure (when measuring the expansion coefficient of a gas)? Were the specific heats of gases measured by reference to a constant weight or a constant volume of gas, and were such 'volumetric' specific heats then measured at constant volume or at constant pressure? Were the conditions diathermal or adiabatic, and was the gas simply 'rehoused' or did it expand against a piston, against the external air, or expand freely against a weightless piston 'against' a vacuum? How good were the vacuums? Was the expansion fast or slow, did the piston slide freely, what was the heat capacity of the vessels, and what was the time-lag of the thermometers? Also, has an isothermal expansion ever been carried out in practice? How is a water-bath maintained at constant temperature and is the specific heat of the water a constant at all the relevant temperatures?

The technology required was also increasingly sophisticated: accurate glassware, brass vessels to sustain high pressures, good-quality pumps and seals, the collection and storage of gases, and so on.

All the while, the gas molecules themselves were still only conjecture.

As regards theoretical advances, this chapter has been long and difficult, but it will be sufficient to hold on to just two findings. First, the temperature T is *the* defining thermodynamic parameter for a gas. Secondly, when a gas does work in expanding or contracting against a piston, then heat is added to or subtracted from the latent heat-reserve in the gas (caloric theory) or is *converted* to or from the work (modern theory of energy).

That there could be such a *conversion* between the mechanical energy of the piston and the heat energy of the gas was not dreamt of in anyone's philosophy at the time. All, that is, except for one philosopher—and that was Dulong. In a letter to Berzelius in January 1820, he wrote:

> I can show that, in making the volume of a gas or vapour vary suddenly, one produces changes in temperature incomparably greater than those which result from quantities of heat developed or absorbed, if there wasn't heat *generated* by the movement. Rumford had already used more or less this line of reasoning…[12]

If Dulong's significant *aperçu* had been noticed, then it might have speeded up the course of physics by 25 years—but no response to this letter has ever been found. We can console ourselves with the knowledge that Dulong was, at any rate, a very nice man. He was a doctor who treated his poor patients in Paris for free and even paid for their prescriptions out of his own pocket. This was despite the fact that Dulong himself always struggled financially, had four children to raise, and suffered from chronically ill health as well as his work-related injuries. He died in 1838, aged 53.

All the progress in empirical knowledge described in this chapter—the gas laws, the specific heats of gases, and the determination of γ—provided the crucial evidence to take Dulong's vision further. Two men in particular, Sadi Carnot and Robert Julius Mayer, were to make great use of the data on the specific heats of gases. Their work (see Chapters 12 and 14) would lead to the formulation of thermodynamics (Carnot) and, ultimately, to the concept of energy (Mayer).

We started this chapter with a look at how the physics of gases helped to give birth to two scientific revolutions, the atomic theory and the theory of energy. The atomic theory was at a more advanced stage in the opening decades of the nineteenth century, but the energy theory had leap-frogged ahead by mid-century. The main reason for the changing fortunes of the atomic theory was philosophical doubts about the reality of atoms. The positivist philosophy of Auguste Comte (1798–1857) was then at its height. (Comte was proclaiming the impossibility of ever knowing the composition of stars even as experimentalists were examining the first stellar spectra.)

It is puzzling that this positivist outlook should not also have hindered the idea of energy. After all, it's not as if anybody actually sees a blob of '$\frac{1}{2}mv^2$' coming towards them. As in Feynman's allegory, the 'blocks' of energy would only emerge through quantitative formulation, experimental corroboration, and the honing of the idea of energy itself.

11

Two Contrasting Characters: Fourier and Herapath

We have already made comment on the different approach to science in England and France during the post-Revolutionary period and we have seen the contrasts between contemporaries such as Watt and Lavoisier (Chapter 8) and Dalton and Gay-Lussac (Chapter 10). Now, once again in 1812–22, two characters appear who epitomize their respective national characteristics: the French professional mathematical physicist, Joseph Fourier, and the English eccentric amateur, John Herapath.

Joseph Fourier

Joseph Fourier (1768–1830) came from humble origins but through sheer talent managed to rise in the fields of mathematical physics, government office, and literary accomplishment. The son of a poor tailor from Auxerre, he was orphaned at eight and would have been lost to science but for a lady who noticed his gentle manners and gentle disposition.[1] She recommended him to the Bishop of Auxerre, who arranged for Fourier to be brought up by Benedictine monks. He soon developed an interest in mathematics and collected candle-ends by day so that he could secretly pursue his studies by night.

When applying for permission to join the artillery, Fourier received the reply 'Fourier, not being of noble birth, cannot enter the artillery, not even if he is a second Newton.'[2] However, after playing an active part in the French Revolution in Auxerre, Fourier was rewarded by an appointment, first at the École Normale, and then at the École Polytechnique. (During the Terror, Fourier's gentle manners were noticed again and he was criticized for being too lenient—and even arrested for his part in helping an innocent man escape the guillotine.[3])

It was while teaching at the École Polytechnique that Fourier's abilities were spotted by Napoleon, and he was asked to accompany Napoleon on his Egyptian campaign in 1798. After three years, Fourier was virtually governor of half of Egypt. Later, Fourier's introduction to the 'Description of Egypt' was so well written that he was admitted to the prestigious French Academy. Upon Fourier's return from Egypt in 1802, Napoleon appointed him Prefect of Isère and he remained there, in Grenoble, for the next 13 years.

It was while he was Prefect of Isère that Fourier first became interested in the subject of heat. Like Boerhaave (Chapter 5) and Lavoisier (Chapter 8), Fourier attributed heat with truly cosmic significance:

> Heat, like gravity, penetrates every substance of the universe, its rays occupy all parts of space.[4]
>
> it influences the processes of the arts, and occurs in all the phenomena of the universe.[5]

Newton's Law of Gravity and Laws of Motion explained both the motions of the heavenly bodies and Earth-bound motions. Likewise, the varied phenomena of heat should be amenable to simple mathematical laws and should apply equally in the celestial and terrestrial domains. Fourier sought to explain the equilibrium that exists between the mean temperature of the planets, the heat radiated by the Sun, and the cold of outer space. He was the first to appreciate the influence of the atmosphere on the surface temperature of the Earth (what we now call the greenhouse effect). He also asked questions such as the following:

> How shall we be able to determine that constant value of the temperature of space, and deduce from it the temperature which belongs to each planet?
>
> What time must have elapsed before the climates [on Earth] could acquire the different temperatures which they now maintain?
>
> From what characteristic can we ascertain that the Earth has not entirely lost its original heat?[6]

In keeping with this cosmic scale, Fourier, despite his claims to a fully comprehensive treatment of heat, was really only interested in the *transmission* of heat, especially by conduction (through the body of a planet) or radiation (across the empty space between the Sun and the planets).

Edmond Halley, at the end of the seventeenth century (Chapter 4), had already demonstrated the 'sine law' for the intensity of radiation emitted from a surface, but Fourier showed that this was a *necessary* law. In a remarkable piece of reasoning, Fourier argued that if the sine law didn't apply then perpetual motion would follow:

> Bodies would change temperature in changing position. Liquids would acquire different densities in different parts, not remaining in equilibrium in a place of uniform temperature; they would be in perpetual motion.[7]

This was probably the first time that the impossibility of perpetual motion had been used in a thermodynamic setting.[8]

Fourier's magnum opus was *The Analytical Theory of Heat* of 1822, in which he was not backward in coming forward, claiming to have 'demonstrated all the principles of the theory of heat and solved all the fundamental problems'.[9] The work was based on his earlier (1812) prize-winning entry for a competition on the modes of transmission of heat.

Jean-Baptiste Biot had carried out experiments (in 1804) in which an iron bar was held hot at one end, cooled at the other end, and left until a steady state was achieved. This meant that at any position along the bar the rate of heat loss by radiation through the surface and by conduction along the bar was exactly matched by heat supplied from the hot end. Biot assumed that the rate of heat loss at the surface was proportional to the excess temperature (i.e. he used Newton's Law of Cooling; see Chapter 5).

Fourier's masterstroke was to mathematically model Biot's experimental set-up in the right (and totally modern) way, by defining the 'flux'—the amount of heat crossing unit area per unit time. His main physical assumption was that Newton's Law of Cooling, of a sort, applied for heat transmission *within* the bar as well as for heat losses from the surface. (In other words, the heat-flux at any distance, x, along the bar was proportional to the temperature gradient, $\partial T/\partial x$, at that distance.) Finally, from the requirements of continuity of heat-flux and conservation of heat, Fourier arrived at his famous heat conduction equation, which is still used today:

$$\partial^2 T/\partial x^2=(c/\kappa)\partial T/\partial t \tag{11.1}$$

where T is the temperature, c is the specific heat capacity, κ is the coefficient of conductivity, x is the distance along the bar, and t is the time.

This equation is curious in a number of ways. First, the temperature varies only to first order in time (i.e. the right-hand side has $\partial/\partial t$ rather than the usual $\partial^2/\partial t^2$ which occurs in so many physics problems, such as the ubiquitous wave equation). This occurs because Fourier's equation is describing the flow of a *massless* fluid ('*calorique*'). But—you may be wondering—how can Equation 11.1 be modern and correct and yet derived from the flow properties of a fictitious weightless fluid? The answer is that, as found in Chapter 10, the right physics can come through even while the metaphor is wrong.

The second strange feature of Equation 11.1 is that the solutions show that some heat is transmitted at infinite speed—absurd and at odds with the fact that heat conduction is a rather slow process. However, the physicality is saved as the *proportion* of heat so transmitted is vanishingly small. (In fact, curiouser and curiouser, it tails off exponentially, in similar fashion to the Maxwell–Boltzmann distribution curve, to be explained in Chapter 17.)

Finally, the main curiosity about Equation 11.1 is that, for the first time in physics, an equation has arisen in which the solution is different if time is reversed (i.e. if $-t$ is substituted for t). There is no evidence that Fourier or anyone else noticed this at the time.

Apart from its correctly describing heat conduction, Equation 11.1 was enormously important for physics, as it led Fourier on to one of the greatest discoveries in mathematics—the Fourier series.

Despite the comment at the end of the previous chapter, the influence of positivism wasn't all negative. Fourier was a positivist and so didn't commit himself to the caloric theory (he encountered some enmity from Biot and Poisson for this, but Laplace himself remained aloof[10]). He always insisted that every mathematical statement of a physical law had to have physical meaning and be capable of measurement. He therefore examined the clumps of physical constants occurring in the exponents of his Fourier series and realized that they had to correspond to something real—the actual 'dimensions' of the physical problem. He proceeded to carry out the first dimensional analysis and to develop a theory of units. This was a major advance, possibly the first since Galileo,[11] in the mathematical representation of physical quantities. Comte adopted Fourier as the leading promoter of positivism in the physical sciences.

Napoleon's downfall also marked the nadir of Fourier's fortunes. The coming of the Restoration and the end of the Hundred Days left Fourier deprived of the Prefecture, in disgrace, and almost destitute. However, the Prefect of Paris, a friend and former pupil, managed to secure for Fourier the directorship of the Bureau of Statistics in Paris. Fourier retained this position until he died in 1830. By a strange irony, Fourier's death was due in part to over-heating—he had a habit of wrapping himself up 'like an Egyptian Mummy' and living in airless rooms at an excessively high temperature.[12] There is some suspicion that he suffered from an under-active thyroid gland.

John Herapath

Every bit the English eccentric, John Herapath (1790–1868) was largely self-educated and was acquainted with the major works of mathematical physics from Newton's *Principia* onwards, and with the recent experimental investigations of Dalton and Gay-Lussac (Herapath had also taught himself French). His style was verbose and opaque, full of made-up words such as 'Numeratom', 'Voluminatom', and 'Megethmerin',[13] set out in scholia, propositions, and lemmas and using proportions instead of calculus.

Yet Herapath, working outside the mainstream scientific community, was one of the founders of the kinetic theory of gases. In fact, there seems to be a correlation between these two groups. First, there was Daniel Bernoulli in Switzerland in the first half of the eighteenth century (see Chapters 4 and 7), 50 years later came Cavendish (see Chapter 8), and then came Herapath in 1816 and John James Waterston in 1845. All worked in isolation (in Cavendish's case, he also worked in secret) and all were unaware of the work of their predecessors. The sad neglect of Waterston's work we shall come to later (Chapter 14).

Like Fourier, Herapath came to heat studies via gravity. He sought to provide a mechanistic explanation of gravity and proposed that a subtle ether, made up from 'gravific' particles, became rarefied by the high temperatures near to celestial bodies. The reduced density of the ether allowed gravity to hold sway. All this brought him to the connection between temperature and particle velocity and to his kinetic theory of gases.

Herapath rejected repulsion as the cause of elasticity in a gas and hit upon the true kinetic theory wherein particles move freely of each other and are in 'projectile motion' as opposed to some sort of vibratory motion. However, he soon came across an obstacle that baffled him and 'was a shock I had hardly philosophy enough to withstand'.[14] This was the inconvenient truth that his particles had to be both perfectly hard and undergo only perfectly elastic collisions. He eventually resolved this dilemma (to *his* satisfaction, at any rate) by making mv rather than mv^2 the fundamental measure of motion. This led him to Boyle's Law and to the correct association between gas pressure and particle speed ($P \propto v^2$), but also to the incorrect $T \propto v$ and $PV \propto T^2$.

Despite the 'unbeatenness of the track'[15] he followed, Herapath was disappointed that his ideas were not better received—even by Humphry Davy, who had supported Rumford in his stand against the caloric theory (see Chapter 9).

Although Davy accepted the idea of heat as motion of the constituent particles, he thought more in terms of a localized vibration or rotation rather than the more mechanistic picture of a gas of randomly moving, non-interacting particles. This may have been due to his affiliation to the 'Romantic' movement in England and France ('*Naturphilosophie*' in Germany) through the influence of his friends, chiefly the poet Samuel Taylor Coleridge. This new Romantic movement, as applied to physics, promoted the idea of mutual influences, ebbs and flows, and quasi-organic evolutions rather than austere mechanism.

As well as this, Davy may simply have found Herapath's style hard to take and his archaic mathematical proofs impossible to follow. There were also some glaring mistakes in Herapath's work.

In any event, Davy informed Herapath that his paper would not be published in the *Philosophical Transactions* (Davy had recently become the President of the Royal Society and the *Philosophical Transactions* was its journal). He advised Herapath to withdraw it, otherwise the paper became the property of the Royal Society and could not be returned to the author. Herapath sent his paper instead to the *Annals of Philosophy*, where it was published in 1821. It was almost completely ignored, although a certain anonymous character, 'X', made some pertinent criticisms in a short reply.[16]

Five years later, Herapath attacked Davy and the Royal Society through the letters pages of *The Times* newspaper. Davy never responded but, when he resigned from the presidency of the Royal Society in 1827, Herapath took it as a victory for himself.

Herapath then retired from the fray. Soon the new era of railways provided him with an opening and in 1835 he became editor of his own *Railway Magazine* (he also had 11 children to support). This gave him a forum to continue publishing his scientific ideas; for example, in 1836 he presented a calculation of the speed of sound that included a calculation of the average speed of a molecule in a gas. This was the first time that this had been done (priority is usually wrongly given to Joule).[17]

Herapath's work went unnoticed until Joule, in 1848, was trawling the literature to find some support for his radical new departure in physics (see Chapter 14). Later still, James Clerk Maxwell acknowledged Herapath's pioneering work. At least Herapath, unlike Avogadro, lived long enough to see the vindication of his ideas with the publication, in 1860, of Maxwell's famous kinetic theory of gases.

Overview

Fourier's heat conduction equation would be of prime importance to William Thomson in his meshing together of Carnot's work and thermodynamics (see Chapter 16). Herapath's work, as we have seen, was eventually recognized by Joule and Thomson.

Another contemporary of Fourier and Herapath, the 'French Faraday', André-Marie Ampère (1775–1836), had yet another approach to heat: he considered its propagation as a wave in his paper of 1832.

However, none of these three approaches to heat—its transmission via conduction or as a wave, or the kinetic theory—considered the connection between heat and *work*. This was all the more surprising as this was the heyday of the steam engine: all over Europe and the East Coast of North America, but especially in Britain, steam engines were pumping water out of mines, powering industry, and transporting people and goods in trains and on boats. These were all examples of the performance of work from heat. It was an engineer rather than a physicist who finally investigated this connection and, in so doing, brought in an overarching new principle in science and initiated the discipline of thermodynamics. We shall cover this in the next chapter.

12

Sadi Carnot

Often, the greatest leap forward occurs with a new approach (for example, when the principle of the economy of nature superseded the principle of the conservation of nature's resources—see Chapter 7, Part IV) or when totally different fields of enquiry are found to have something in common (as when Newton embraced the motion of the Moon and the falling of an apple in a single set of laws).

Contemporaneous with the French Laplacian school (Laplace, Biot, and Poisson) and also Ampère, Dulong, Fourier, and others, came a revolutionary new departure. It was, though, a very quiet revolution, as we shall see. All the workers listed above were physicists, whereas the new departure came from a young engineer, Sadi Carnot (1796–1832). Contrary to common perception, the influence from engineering to physics is rare; the spin-off usually goes the other way.

Laplace and his followers tried to understand the forces between heat-particles (caloric) and matter-particles. Fourier asked what laws of heat transmission were necessary to ensure equilibrium in the motion of celestial bodies. But Carnot was the first to ask about the connection between heat and motion on an everyday, engineering scale—not motion on a microscopic or astronomical scale, but the sort of motion that occurs in the workings of a heat-engine.

Sadi Carnot asked: What, if any, are the limits to the efficiency of a heat-engine?

We can see in this question a concern with engineering, but also with social issues such as economics. On both counts, Sadi's chief influence was his father, Lazare Carnot, a mathematician and engineer, and a leading statesman during the time of the French Revolution (see Chapter 8).

Lazare named his son Sadi after a medieval Persian poet and moralist, Saadi Musharif ed Din. There are not many anecdotes from Sadi's short life, but all show him to be a person with high moral standards. For example, when the Buonaparte and Carnot families were on an outing, and Napoleon was throwing stones to splash his wife and other ladies in a boat on a lake, Sadi, aged four, ran up and shook his fist at Napoleon, shouting: 'You beastly First Consul, stop teasing those ladies!'[1] (Fortunately for physics, Napoleon just laughed.) Later, as a young adult, Sadi drew up a list of rules of conduct for himself: 'Say little about what you know and nothing at all about what you don't know... When a discussion degenerates into a dispute, keep silent... Do not do anything which the whole world cannot know about...'[2]

Coming from a famous, political family, Sadi was motivated to improve the economic and political standing of France. He understood the huge importance of the heat-engine: 'The study of these engines is of the greatest interest, their importance is enormous, their use is continually increasing, and they seem destined to produce a great revolution in the civilized world.'[3] He also understood the particular importance of the steam engine to England:

> To take away today from England her steam-engines would be to take away at the same time her coal and iron. It would be to dry up all her sources of wealth, to ruin all on which her prosperity depends, in short, to annihilate that colossal power. The destruction of her navy, which she considers her strongest defence, would perhaps be less fatal.[4]

Now steam engines had been improved steadily from the time of Savery onwards and Sadi acknowledged the enormous debt of the world to the eighteenth-century English engineers Newcomen, Smeaton, 'the famous Watt', Woolf, and Trevithick. Sadi wanted to know whether the work that could be obtained from a heat-engine, the 'motive power of heat', was unbounded or whether there was an assignable limit to the improvements that could be made: 'a limit which the nature of things will not allow to be passed by any means whatever'.[5]

The laws of operation of machines not using heat (the simple machines such as the lever, pulley, and inclined plane and more complicated machines powered by water, wind, or muscle) were all well known and ultimately attributable to the mechanical theory (noted Carnot). Sadi's own father, Lazare, had initiated the fundamental theory of such machines.

Sadi's genius lay in his isolation of the one law relevant to all these examples and his realization that this law would apply to *all* machines, whether mechanical or *heat-driven*. He set out a simple argument, using no mathematics and couched in the style of the ancient Greeks—the syllogism. His very argument was ancient—he appealed to the age-old law of the impossibility of perpetual motion.

His argument was: hypothesize an ideal engine that is the best possible—it yields the maximum amount of work. It is also reversible. Now imagine that there exists another engine that could produce *more* work than this ideal engine, say, the 'superior' engine. If the ideal engine was run in reverse, then the superior engine could totally counteract it—it could replace all the work consumed and return the world to the identical starting conditions—but there would still be some work-capacity left over. This leftover work could be used to generate a perpetual motion. However, perpetual motion is not possible and therefore the initial premise (that a superior engine exists) is ruled out.

In succinct syllogistic form the argument runs as follows: perpetual motion is not possible; a superior engine would permit perpetual motion; therefore a superior engine is impossible. In other words, there can be no engine better than an ideal engine.

So far, this result is not too surprising. After all, we hypothesized the ideal engine as the best possible one in the first place. However, more can be teased out of the argument. There can be no engine better than an ideal engine, but can there be a worse one? Obviously, in the case of ordinary engines (real, actual engines), these are all worse than the ideal engine, but could one ideal engine be worse than another? For example, suppose that ideal engine IEworse is worse than ideal engine IE. Then IE would be better than IEworse. But our original argument prohibits this—no engine, ordinary or ideal, can be better than the ideal engine.

So, combining the two results, no ideal engine can be better or worse than another ideal engine. We finally come to the remarkable conclusion: all ideal engines perform equally well, and this will be true whatever the details of their construction, whether they use this fuel or that, and so on.

We will analyse this argument.

First, we must be sure that we are only comparing ideal engines of the same 'size' even while they might work in different ways. For example, if we were to employ IE ten times over, then it would be equivalent to one ideal engine ten times larger and would obviously do ten times as much work. We shall define the 'size' of an ideal engine a bit later on.

Secondly, we note that Sadi was careful to define 'perpetual motion' and show that it could be of two kinds. It was 'not only a motion susceptible of indefinitely continuing itself after a first impulse [has been] received, but [also] the action of an apparatus, of any construction whatever, capable of creating motive power in unlimited quantity'.[6] This was a subtlety that had already been appreciated by Leibniz (see Chapter 3). By ruling out *both* kinds of perpetual motion, Carnot was bringing in what would become the Second Law of Thermodynamics as well as what would become the Law of Conservation of Energy and the First Law of Thermodynamics.

We also note that Sadi thought to justify his premise of the impossibility of perpetual motion. As we saw in Chapter 2, this is one of those truths that has always been challenged even while it has always been accepted. Sadi says:

> The objection may perhaps be raised here, that perpetual motion, demonstrated to be impossible by mechanical action alone, may possibly not be so if the power either of heat or of electricity be exerted; but is it possible to conceive the phenomena of heat and electricity as due to anything else than some kind of motion of the body, and as such should they not be subjected to the general laws of mechanics?[7]

This didn't mean that Sadi accepted the motion theory of heat—in fact, he adopted the mainstream caloric theory—but that he assumed that, at the most fundamental level, all processes were mechanical. Sadi was also convinced by the fact that, experimentally, *all* attempts at perpetual motion had failed: 'Do we not know besides, *a posteriori*, that all the attempts made to produce perpetual motion, by any means whatever, have been fruitless?'[8]

Finally, we note that Sadi's argument required an ideal engine to be reversible.

While Sadi's argument is admirable, our admiration is tempered by the fact that we want to know what *is* an ideal heat-engine?

Sadi was influenced, of course, by his father's seminal work on the efficiency of mechanical engines (see Chapter 8, Part I) and he wanted to find similar strictures for heat-engines. Lazare had specified conditions for optimizing the performance of any machine in his Principle of Continuity. By this Principle, power should be transmitted smoothly and continuously, without percussion or turbulence.

Take, for example, the case of a water-engine driven by a fall of water, the scenario so often considered by Lazare. We can imagine two waterwheels, A and B, and, once again, use the argument of the impossibility of perpetual motion to state that B can't generate more work than A; otherwise, B could be run in reverse, employed as a 'water-raiser', and be used to drive A, perpetually. But this doesn't clinch the proof—perhaps B has better-shaped scoops and does, in fact, work more efficiently than A.

For a complete proof, we must first idealize the water-wheel, A, to be the best possible—that is, *ideal*—water-wheel. The falling water must ideally meet the scoops with zero relative velocity so that there are no losses due to 'shock';[9] there must be no loss of water due to spray or splashes; no friction at the bearings; the water must trickle out at the bottom with zero speed; and so on. Only then can we use it in our argument and assert that, yes, this idealization *is* the best possible, no water-wheel could work better, and so a perpetually acting machine is ruled out.

Behind the details of this idealization lies the deeper understanding that it is the falling of weights (the weight of water falling through a given height) that is the source, sole determinant, and measure of the work done by any water-engine working under gravity.

Now, getting back to Sadi, we ask: What is the deeper understanding behind the motive power of heat? What is the idealization in the case of a *heat*-engine? It is here that Sadi Carnot's genius truly came to the fore. He understood, for the first time, the quintessential nature of a heat-engine: *it is a machine that does work as a result of heat flowing between two temperatures*.

It is hard to be impressed enough by the simplicity of this, 200 years later, when the physics is all well known. However, the sheer variety of heat-engines around in Carnot's time persuades us of his achievement. There were steam-driven paddle-wheels, contemporary versions of Amontons' fire-wheel (see Chapter 4), Newcomen's atmospheric engine (see Chapter 5), Watt's condensing engine (see Chapter 8), Trevithick's high-pressure steam engines, compound engines, engines using the expansive properties of air or alcohol vapour, combustion engines, and buoyancy engines. We could also include heat-engines in the more abstracted sense, such as the cannon and the hot-air balloon, or take examples from geology (e.g. volcanoes) or meteorology (e.g. the water-cycle). Carnot saw that in *all* these 'engines' the overriding feature was that a certain quantity of heat flowed between two temperatures. But—which two temperatures?

Another crucial, fundamental, and simple understanding came in with Carnot: heat always flows from a high temperature to a lower temperature. This knowledge was as old as the hills, but who had thought to raise awareness of it and bring it into physics? Even Black (see Chapter 6), who understood that separate bodies will lose or gain heat until they achieve an equilibrium of temperature—even he did not bring out the fact that heat always flows from hot to cold. With Carnot, *direction* (of a process) came into physics for the first time.

Let's get back to the idealized heat-engine and see how Carnot makes it approachable in reality. After all, a water-wheel may be nothing but a machine driven by a flow of water between two heights, but a real water-wheel must have the best design of scoops, the most friction-free bearings, and so on.

Carnot's realization was that when heat flows it can do work by causing a change in volume. For example, a gas could be heated, made to expand, and thereby push a piston that could be linked to a weight-raising machine. He did see the possibilities of other sorts of action, such as chemical, but chose to concentrate on the most commonplace effect of volume change. He had studied actual heat-engines in use across Europe, mostly driven by steam, and was aware of the special connection between gases (or vapours) and heat, and of all the contemporary research into the physics of gases (see Chapter 10).

Carnot stressed that any heat-flow that occurs but does not result in a volume change is wasted—the potential for doing work was there but has not been exploited. Therefore, in an ideal heat-engine, there should be no direct (i.e. thermal) contact between bodies at different temperatures.

The final feature of Carnot's ideal heat-engine was that it should be reversible—his syllogistic argument requires this feature.

We therefore have the requirements that an ideal heat-engine should operate by:

(1) A heat-flow between two temperatures.
(2) No *direct* heat-flow between these temperatures (i.e. no contact between bodies at different temperatures).
(3) only a reversible heat-flow between these temperatures.

These multiple requirements appear incompatible, especially when combined with the fact that heat only flows where there is a temperature gradient and then only in one direction, from hot to cold.

Carnot's remedy is novel, simple, abstract, and ingenious—and, furthermore, it will become the basis of the whole approach of modern thermodynamics. In a nutshell, it is that, in an ideal heat-engine every temperature change, every heat-flow, shall occur only by infinitesimally small steps. Therefore each such step will be individually reversible.

Direct contact between bodies at different temperatures *is* allowed, but only where the temperature difference is infinitesimal. In fact, such temperature differences must be allowed, otherwise no heat will flow. Heat will still only flow from hot to cold, but the temperature gradient, being infinitesimal, can always be reversed at will.

We have already acknowledged the great influence of Lazare Carnot on his son. Lazare inspired Sadi to consider the utility of machines for the glory of France and for all mankind: 'If real mathematicians were to take up economics and apply experimental methods, a new science would be created—a science which would only need to be animated by the love of humanity in order to transform government.' Lazare's work on the theory of generalized machines and his Principle of Continuity were undoubtedly starting points for Sadi. However, perhaps the greatest legacy of Lazare to Sadi was the concept of reversibility.

Lazare had introduced the idea of infinitesimal 'geometric motions' (see Chapter 8) and further stipulated that these were reversible ('the contrary motions were always possible'[10]). Sadi incorporated this idea of infinitesimal reversibility but added a brilliant extension to it: for a succession of individually reversible, infinitesimal processes, one could arrive at a *macroscopic* process that was reversible—the idealized heat-engine.

All this is very abstract. We turn now to Sadi's own description of how such an ideal heat-engine would run. This comes straight from Sadi's historic book, *Reflections on the Motive Power of Fire*,[11] a work embarked upon after a visit to his father[12] in exile in Magdeburg in 1821. Sadi never saw his father again and the book was published in 1824, one year after Lazare's death. Sadi first sets the scene (Fig. 12.1 is from his work):

> let us imagine an elastic fluid, atmospheric air for example, shut up in a cylindrical vessel, abcd (Fig. 1) [Fig. 12.1], provided with a moveable diaphragm or piston, cd. Let there be also two bodies, A and B,* kept each at a constant temperature, that of A being higher than that of B.[13]

Sadi takes the starting position of the piston to be ef and the cylinder to be initially isolated from the heat reservoirs A or B. He describes the following cyclic series of operations (his numbering has been left intact):

(3) . . . The piston . . . passes from the position ef to the position gh. The air is rarefied without receiving caloric, and its temperature falls. Let

* A and B are reservoirs of heat at temperatures T_A and T_B, respectively.

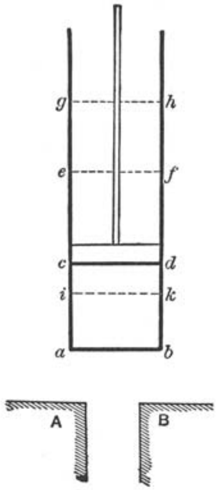

Fig. 12.1 Carnot's diagram of his cycle, in *Reflections on the Motive Power of Fire*, 1824 (by permission of Dover Publications, Inc.).

us imagine that it falls thus till it becomes equal to that of the body B; at this instant the piston stops, remaining at the position gh.

(4) The air is placed in contact with the body B; it is compressed by the return of the piston as it is moved from the position gh to the position cd. This air remains, however, at a constant temperature because of its contact with the body B, to which it yields its caloric.

(5) The body B is removed, and the compression of the air is continued, which being then isolated, its temperature rises. The compression is continued till the air acquires the temperature of the body A. The piston passes during this time from the position cd to the position ik.

(6) The air is again placed in contact with the body A. The piston returns from the position ik to the position ef; the temperature remains unchanged.

(7) The step described under number (3) is renewed, then successively the steps (4), (5), (6), (3), (4), (5), (6) . . . and so on.[14]

This is the famous Carnot cycle. It is most easily understood by means of a pressure against volume graph, or *PV* diagram (see Chapter 10).

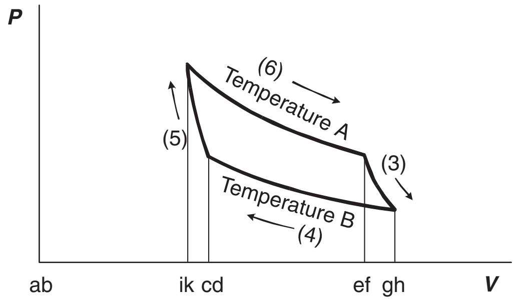

Fig. 12.2 Carnot's cycle shown on a graph of pressure versus volume.

(Carnot himself never used such an 'indicator diagram'. It was used to describe Carnot cycles for the first time by Émile Clapeyron in 1834; see Chapter 16.)

The path from ef through gh, cd, ik, and back to ef again, roughly a parallelogram, traces the various 'states' (pairs of P and V values) of the air enclosed in the cylinder as it expands or is compressed. As we know from Chapter 10, the (uncompensated) air temperature rises during the compressions and falls during the expansions. (This latter is because none of the expansions are 'free'; that is, the expanding air pushes back on the surrounding external air and does work all the while—see Chapter 10). The whole cycle constitutes the ideal heat-engine. (We can see the influence of Watt's steam engine: section (3) is the same as Watt's 'expansive operation', except that in the latter case the amount of steam is continuously throttled down.)

There are many qualifications that must be added to this sparse outline. We are to understand that the 'bodies' A and B are infinite reservoirs of heat. They stay at the temperatures T_A or T_B however much heat is added to or taken away from them. We are also to assume that the piston moves without friction or leaks; the heat transmission between cylinder and reservoir is perfect; and the thermal isolation of the cylinder, when not in contact with a reservoir, is total.

Moreover, all the operations must be reversible and must therefore be carried out smoothly and gradually. In effect, Lazare's Principle of Continuity must be adhered to for heat-engines as well as for mechanical engines. When the air is compressed, it is understood that this is to be achieved by applying an external force to the piston. However, the pressure so applied must be increased slowly and continuously so that it is always only marginally above the pressure inside the cylinder. Likewise, for expansions, the pressure outside the cylinder must be continually adjusted so that it is at all times only marginally less than the internal pressure. Sadi doesn't explicitly mention these provisos. He also tacitly assumes that the temperature and pressure can be monitored continuously without compromising their values in any way.

With these idealizations, we see that during the sections ik to ef and gh to cd (i.e. when the cylinder makes contact with the given heat reservoir) the air in the cylinder remains at the constant temperature T_A or T_B, respectively. These sections are therefore isothermal and follow the Ideal Gas Law, $PV = nkT$ (see Chapter 10). (Sadi would have described this as Mariotte's Law.)

In the sections ef to gh and cd to ik, the cylinder does not lose or gain heat from the surroundings and the temperature therefore changes by a finite amount, but continuously and slowly as the expansion or compression proceeds. The internal air pressure also changes continuously and for *two* reasons: the temperature change just mentioned and the volume change. The path followed is thus the steeper 'adiabatic' curve (see Chapter 10). Sadi knew that the curve was steeper but not that it followed PV^{γ} = constant, where $\gamma = C_P/C_V$, although this relationship had already been worked out by Poisson a year or two earlier.

We now ask how much work this ideal engine does. We could gear up the piston to raise and lower weights; however, we already know the work output for the volume change of an ideal gas. It is $\int P\,dV$, the area under the PV curve (see Chapter 8, Part I, the section on Davies 'Giddy'). The net work is given by the area enclosed by the PV closed curve or cycle. As we have the convention that the engine *does* work during an expansion, it then follows that net work is done in a *clockwise* cycle.

Also in this clockwise direction, and during the isothermal sections, heat will be transferred continuously *from* the heat reservoir A (to the expanding and continuously cooling air in the cylinder) and *to* the heat reservoir B (from the continuously compressed air). Therefore, when the engine is run in the direction so as to produce net work, there will also be a net flow of heat from the higher temperature, T_A, to the lower temperature, T_B.

We thus have a heat-engine that conforms to Sadi's overall specification: work is done and heat flows from a higher to a lower temperature. Moreover, the engine is reversible; that is, the exact same path can be tracked but in a reverse (anticlockwise) direction. In this reverse direction, the engine is a net consumer of work and there is a net transfer of heat from B to A. But Carnot's directive, that heat flows only from hot to cold, has not been overturned. Even in the idealized heat-engine, and even when it runs 'backwards', the heat still never flows *directly* from cold to hot. It trickles from the cold reservoir B to the even cooler, expanding gas; and from the hot, compressed gas to the infinitesimally cooler A.

The proof of the reversibility of the engine for Sadi was the fact that there is a return to the initial 'state' of the air. For example, at ef (Fig. 12.1), the air has exactly the same P and V values however many times this state has been passed through and from whichever direction it is approached. This idea of the *state* of a system, uniquely defined by certain macroscopic parameters—say, P and V—is the second great gift of Sadi Carnot to thermodynamics (after the idea of reversibility). His third great gift was the idea of a complete cycle: for any arbitrary starting point, the ideal engine must run through a complete series of operations (i.e. trace a *closed* path on the indicator diagram) and end up where it began.

The importance of all this was so that different ideal engines could be compared. Only by carrying out a complete cycle and returning to the same starting state could Carnot be sure that there were no work costs or heat costs unaccounted for. One engine could then be compared to another and the efficiency of engines gauged. We remember that improving the efficiency of engines was Sadi's ultimate goal.

We remember also Sadi's remarkable proof that all ideal heat-engines have exactly the *same* efficiency, regardless of the type of engine, choice of working substance, details of construction, and so on. This, in fact, is the reason why we have been able to employ the specific case of the PV cycle (as opposed to, say, electrical or chemical operations) to stand in for the abstract Carnot engine. This is not a cheat. Air at typical densities and temperatures approximates very well to an ideal gas (as defined in Chapter 10), and with all the other provisos (large heat reservoirs, slow operations, and only tiny differences in temperature) the air-in-a-cylinder engine does represent almost exactly an ideal case.

So, now that this ideal heat-engine has been defined, we ask: What did Sadi find? What *is* the ideal efficiency?

We must be sure that we know what 'efficiency' means. Surely, an engine using more heat will yield more work. And an engine with a greater drop in temperature will also do more work. Sadi took his clue from the analogy with water-power (like his father Lazare, Sadi always had water-engines acting by gravity in mind):

> we can compare with sufficient accuracy the motive power of heat to that of a waterfall.... The motive power of a waterfall depends on its height and on the quantity of the liquid; the motive power of heat depends also on the quantity of caloric used, and on what may be termed, on what in fact we will call, the height of its fall, that is to say, the difference of temperature of the bodies between which the exchange of caloric is made.[15]

In other words, Sadi understood that the heat-engine will do more work for a greater amount of heat used and for a greater 'fall' in temperature. This answers our earlier question as to the 'size' of engines. Only engines using the same quantity and fall of heat have the same 'size', and only these engines are to be compared.

So, we compare engines that are ideal and of the same size.

Now, for the first time, comes a real difference between water- and heat-engines. Ideal water-engines, whatever their mechanism, all perform at 100% efficiency. However, we find that heat-engines, even ideal ones, never operate with an efficiency of 100%. This inefficiency is *sui generis*, built into the very definition of the heat-engine. For Sadi's heat-engine must operate between *two* temperatures, and heat must always be put out at the lower temperature:

> heat alone is not sufficient to give birth to the impelling power: it is necessary that there should also be cold; without it, the heat would be useless.[16]

We can, in today's terms, see that this difference between the two sorts of ideal engine, heat- and water-, occurs for the following reason. Water flows to the bottom level, but it has been completely stripped of its (gravitational potential) 'energy'. Heat (the subtle fluid, caloric) also flows to the bottom level or temperature, but it cannot be stripped of its energy. In fact, heat *is* energy. Sadi, of course, knew nothing of energy. He could not appreciate that by his insistence that the heat-engine throws out heat at the lower temperature, he was limiting its efficiency.

Perhaps the efficiency of the ideal heat-engine is thus limited to some fixed amount—say, 70%?

We have been careful to compare only engines using the same quantity of heat and fall in temperature, but we have not thought to specify the *actual* temperature. Sadi does not beg any questions:

> We do not know... whether the fall of caloric from 100 to 50 degrees furnishes more or less motive power than the fall of this same caloric from 50 to zero.[17]

This is quite startling. It is like asking whether a waterfall of exactly 50 m height, and for the same quantity of water (say, 100,000 kg), will yield more work at 0 m above sea level than at 1,000 m above sea level.* But this is exactly what Carnot does find in the case of heat-engines:

The fall of caloric produces more motive power at inferior than at superior temperatures.[18]

Specifically, a given quantity of heat falling through a given temperature interval can do more work the lower the temperature. The efficiency is therefore not only less than 100%, it actually *depends* on the temperature. This is the conclusion that will eventually lead to the Second Law of Thermodynamics—that the 'worth' of heat is temperature-dependent.

Carnot couldn't fully justify this conclusion on his own calculations. It was rather his physical intuition that told him that it had to be true. However, this temperature dependence of the ideal efficiency can't have come as a complete surprise to him. After all, he had already (implicitly) conceded a temperature dependence in the very *modus operandi* of his engine (if there was no difference in the work carried out in the low- and high-temperature sections of the cycle, then the engine would yield no net work).

So, how did the efficiency of the ideal heat-engine compare to that of real engines? We still have not defined 'efficiency': for a water-fall engine, it is the work done for a given weight of water falling through a given height; for a heat-engine, it is the work done for a given quantity of heat dropping through a given temperature interval. The work done in each case can be easily determined. But now we come to yet another significant difference between water- and heat-engines. The water can be seen as it flows but heat ('*calorique*') is a weightless and invisible fluid. It cannot be seen as it trickles into or out of heat reservoirs, and these reservoirs, being essentially infinite, leave no tally.

So Carnot had to resort to a blend of contemporary data (the whole corpus of nineteenth-century work on gases; see Chapter 10) and 'exact reasoning'.[19] In detail, he used 'Poisson's result' (that an adiabatic compression of 1/116th leads to a 1°C temperature rise); Gay-Lussac's expansion coefficient (heating a gas directly by 1°C at constant pressure leads to an expansion of 1/267th); the (ideal) gas laws of Mariotte and of

* In fact, as the strength of gravity varies inversely with height, there *is* a difference between the waterfalls.

Gay-Lussac; the measurements of γ; the measurements of specific heats by Dulong and Petit, Delaroche, and Bérard; the work of Dalton, and of Clément and Desormes; and many other results. All these results are given in Chapter 10, but it is necessary for us only to follow the gist.

Using the above data, Carnot tracked the heat and the work done by three hypothetical engines—using air, water vapour, and alcohol vapour. His results were impressive and, in particular, these last two engines had efficiencies agreeing to within 1.5%.

His overall conclusions were as follows:

- The ideal efficiency has a maximum, less than 100%, and is independent of the working substance and design of the ideal engine.
- This efficiency depends only on temperature (it depends on pressure only through this dependence on temperature).
- The efficiency is greater the greater the temperature interval.
- For a given temperature interval the ideal efficiency is probably a continuous function of the temperature ('Carnot's function'), being larger at smaller temperatures.

But Carnot had only been able to calculate *relative* efficiencies. If he had been able to calculate the absolute efficiency of the three hypothetical engines, he may have been less than impressed. For example, for the alcohol-vapour engine running at around 78 °C and with a one-degree drop in temperature, the ideal efficiency is only 0.28% (see the calculation after Equation 16.1 in Chapter 16). This low absolute efficiency reinforces the conclusion of Chapter 8—that getting work from heat is inherently difficult.

Whether or not Carnot suspected this inherent limitation is not known. However, it didn't contradict any of his conclusions and, moreover, didn't undermine his intention to beat the British steam engineers at their own game. The close agreement between the relative efficiencies appeared to vindicate Carnot's initial proposition, which he now elevated to law:

Carnot's Law
The motive power of heat is independent of the agents employed to realize it; its quantity is fixed solely by the temperatures of the bodies between which is effected, finally, the transfer of caloric.[20]

Carnot, in the manuscript version of his book, writes triumphantly:

The fundamental law that we proposed to confirm seems to us to have been placed beyond doubt, both by the reasoning which served to establish it, and by the calculations which have just been made.[21]

However, in the *published* version of the *Réflexions*, his tone is completely different:

The fundamental law that we proposed to confirm seems to us to require, however, in order to be placed beyond doubt, new verifications. It is based upon the theory of heat as it is understood today, and it should be said that this foundation does not appear to be of unquestionable solidity.[22]

What had caused this hesitancy?

As we have argued above, Carnot was not depressed by the low absolute efficiency of the ideal heat-engine, his confidence was eroded for a different reason. It appears that, in the short time interval between the manuscript and published copies, Sadi had the dawning realization that his 'exact reasoning' was flawed in one pivotal respect. He had used the caloric theory of heat and had adopted without question its central tenet (see Chapter 8), that heat is always conserved.

Water fell, the blades of a water-wheel were set a turning, and all the water then arrived at the bottom level, at the end of its fall. And should not the subtle fluid, caloric, likewise 'fall', set a heat-engine to do work, and then all be collected at the bottom level, i.e. the lower temperature? But Sadi's intuition began to tell him otherwise: did *all* the heat really arrive at the lower temperature? Carnot must have begun to see that this was false, and with this realization came a new idea—heat could be *consumed* and *converted* to or from motive power.

After Carnot's death, a bundle of his papers was found—some 23 loose sheets labelled 'Notes on mathematics, physics and other subjects'[23] and thought to span a time from the writing of his book until his death in 1832. Here is a selection from these posthumous notes, in assumed chronological order:[24]

[A]lways in the collision of bodies there occurs a change of temperature, an elevation of temperature. We cannot, as did Berthollet, attribute the heat set free in this case to the reduction of the volume of the body; for when this reduction has reached its limit the liberation of heat should cease. This does not occur.

how can one imagine the forces acting on the molecules, if these are never in contact with one another…? To postulate a subtle fluid in between would only postpone the difficulty, because this fluid would necessarily be composed of molecules.

Is heat the result of a vibratory motion of molecules? If this is so, quantity of heat is simply quantity of motive power.

Can examples be found of the production of motive power without actual consumption of heat? It seems that we may find production of heat with consumption of motive power.

Supposing heat is due to a vibratory movement, how can the passage from the solid or the liquid to the gaseous state be explained?

When motive power is produced by the passage of heat from the body A to the body B, is the quantity of this heat which arrives at B (if it is not the same as that which has been taken from A, if a portion has really been consumed to produce motive power) the same whatever may be the substance employed to realize the motive power?

Is there any way of using more heat in the production of motive power, and of causing less to reach the body B? Could we even utilize it entirely, allowing none to go to the body B? If this were possible, motive power could be created...without consumption of fuel, and by mere destruction of the heat of bodies.

If, as mechanics seems to prove, there cannot be any real creation of motive power, then there cannot be any destruction of this power either—for otherwise all the motive power of the universe would end by being destroyed.

At present, light is generally regarded as the result of a vibratory movement of the ethereal fluid. Light produces heat, or at least accompanies the radiant heat... Radiant heat is therefore a vibratory movement. It would be ridiculous to suppose that it is an emission of matter while the light which accompanies it could only be a movement.

Could a motion (that of radiant heat) produce matter (caloric)?

Undoubtedly no; it can only produce a motion. Heat is then the result of a motion. Then it is plain that it could be produced by the consumption of motive power and that it could produce this power.... But it would be difficult to explain why, in the development of motive power by heat, a cold body is necessary; why motion cannot be produced by consuming the heat of a warm body.

Heat is simply motive power, or rather motion which has changed its form. It is a movement among the particles of bodies. Wherever there is destruction of motive power, there is at the same time production of heat in quantity exactly proportional to the quantity of motive power destroyed. Reciprocally, wherever there is destruction of heat, there is production of motive power.

We can establish the general proposition that motive power is, in quantity, invariable in nature; that it is, correctly speaking, never either produced or destroyed. It is true that it changes its form—that is, it produces sometimes one sort of motion, sometimes another—but it is never annihilated.

In summary, we can see that Carnot's thoughts evolved away from the caloric theory and to the dynamical theory of heat. He came to realize that: heat cannot be material, it must be 'a motion'; all motive power is 'motion' of one sort or another; and that interconversions are possible, but not the creation or destruction of 'motion'.

Carnot also suggested various experiments to test and quantify the conversion between heat and work—experiments that foreshadowed those of Mayer and Joule some 20 years into the future. He estimated the conversion factor between heat and work (what would later come to be known as the 'mechanical equivalent of heat'), and his value is the first to appear in print.

Finally, Carnot's posthumous notes show that he wondered whether or not *all* the heat in a body could be converted into work. This was a question not to be resolved until the Second Law of Thermodynamics had been formulated.

However, Carnot's goal was to improve engine efficiency rather than to discover new laws of physics. He concludes the *Reflections* by giving advice on engine improvements to engineers. Overall, he stipulates that: attending to reversibility in the small will ensure efficiency in the large whatever is the mechanism of the given engine; insulation and avoidance of contact between parts at different temperatures will minimize the 'useless flow of heat'; and, finally, the highest starting temperature and the largest possible fall in temperature is desirable. For this reason, air might be a better working substance than steam.

Despite the contemporary success of high-pressure steam engines, Carnot suspects that the high pressures are actually more of a hindrance than a boon. They can cause cylinders to burst and therefore lessen the maximum temperature that can be safely achieved. Air in a combustion engine would be more suitable than steam in this respect. He had thus anticipated the technology of an age some 40 years into the future (the first internal combustion engine was that of Jean Lenoir in 1859, and by 1885 Karl Benz was driving the first automobile).

Carnot had therefore understood that it is *temperature* rather than pressure that is ultimately the driver of the heat-engine. This is in accordance with our findings in Chapter 10, where we concluded that it was T rather than P that was the thermodynamic parameter of importance.

Overview

Despite Carnot's amazing achievements, he never actively promoted his book. It is true that he wasn't the self-promoting type (by all accounts, he was very introverted). Nevertheless, the main reason for Carnot's reticence must lie with his understanding that the very cornerstone of his theory had crumbled away. He was also reticent about his abandonment of the caloric theory, and never disclosed his changing views, even to his friend and associate Nicolas Clément. The prevailing hegemony was too strong, his views too radical.

Carnot would doubtless have known that his work would stand the test of time, and that it could be reworked to incorporate the dynamic theory of heat—and, of course, this *has* been done (see Chapter 16)—but this was too big a task for one individual to carry out.

The book itself was not a success. It was both too abstract for the engineers and too far from the mainstream for the physicists—in short, it was years ahead of its time. Six hundred copies were printed and cost 3 Francs each (in 1824). One review was given—it was favourable, but then the editor of the journal was Sadi's brother, Hippolyte. Then the book simply disappeared from view.

Apart from a few copies of the *Reflections*, one other research paper, and the posthumous notes already mentioned, all Sadi's work has been lost. He contracted cholera in the great epidemic of 1832 and died within a few hours, aged only 36. All his personal effects (save a picture of a champion boxer)[25] were burned.

The *Reflections* would have been lost to physics altogether but for one thread. The engineer Émile Clapeyron (1799–1864) found a copy, modified it (chiefly by the addition of the indicator diagram) and applied it to the liquid/vapour boundary in 1834.[26] This work was discovered by William Thomson (later, Lord Kelvin) in 1845 and inspired Thomson to look for Carnot's original work. Thomson scoured the booksellers of Paris, who had never heard of the *Reflections*. Eventually (in 1848), Thomson found a copy and pronounced it 'the most important book in science I have ever read'.[27] It was a turning point of his career and of the fortunes of Carnot's work (see Chapter 16).

Carnot had correctly identified the essential method of a heat-engine: heat flows from a high temperature to a lower temperature while doing work (a volume change against an external pressure, a change in chemical configuration, and so on). It is plausible that this crucial link couldn't have been made until James Watt had discovered the condensing steam engine, with its readily apparent temperature change from the hot boiler to the cold condenser (see Chapter 8).

Carnot had found that the ideal efficiency was independent of the method or working substance of the heat-engine (whether gas, vapour, liquid, etc.) and was probably determined solely by the actual temperatures between which the heat flowed. However, the maximum efficiency of even an ideal heat-engine was inherently limited and was lower at higher temperatures (for a given fall of heat).

Finally, Carnot had also set out the practical features of how to maximize the work-output of real heat-engines: heat leaks must be minimized, the temperature of the hot source should be as high as possible, and the temperature of the cold sink as low as possible.

Carnot had therefore accomplished all that he had set out to achieve, and he had unwittingly also set out the groundwork for the whole of thermodynamics. He had introduced the concepts of 'state of a system', 'cycle of operations', and 'reversibility'; defined the ideal heat-engine; recognized that temperature rather than pressure is the primary determinant; and realized that the maximum amount of work to be obtained from heat probably depended on nothing other than the absolute temperature of that heat. This last would eventually lead to the Second Law of Thermodynamics. However, perhaps the most remarkable aspect of Carnot's work is that so much should be right, and of long-term utility to science, when his fundamental axiom, that total heat is conserved, was wrong.

We shall examine this paradox.

To keep track of the heat, there is one telltale which experiment could, in principle, supply (in the case of gases)—the specific heat capacity of the gas *when heated directly*; for example, by warming the gas over a burner. However, as we saw in Chapter 10, the data were difficult to obtain, and, in one case, plain wrong. The latter refers to the data of Delaroche and Bérard, who found that the specific heat of gases increased with an increase in volume. But Carnot's instinct for what was reliable—and *physical*—stayed with him:[28]

> According to the experiments of MM. Delaroche and Bérard, this capacity [the specific heat capacity at the two 'volumes'] varies little—so little even, that differences noticed might strictly have been attributed to errors of observation…
>
> [Although] it is only necessary to have observed them [the specific heats] in two particular cases…the experiments…have been made within too contracted limits for us to expect great exactness in the figures…
>
> Experiments having shown that this heat [the specific heat of air] varies [only] little in spite of the quite considerable changes of volume…If we consider it nothing…[then] *the motive power produced would be found to be exactly proportional to the fall in caloric*.[29] (emphasis added)

Apart from their presumption in using just two data points to derive a whole relationship, Delaroche and Bérard—along with all the physicists of the period—had mistakenly made an equivalence between the *direct* heating of a gas and the *adiabatic* heating that occurs as a gas is compressed or expanded. (This has been explained in Chapter 10.)

Laplace was one who had considered adiabatic changes from a theoretical standpoint—the caloric theory, of course. ('Adiabatic'—see Chapter 10—means that no heat enters or leaves the system: all temperature changes in the gas occur as a result of a volume change against a pressure). Laplace had famously used these adiabatic effects to derive a correction factor, $\sqrt{\gamma}$, for Newton's erroneous determination of the speed of sound in air (see Chapter 10, Part II). Poisson's theoretical results were also derived using Laplace's work.

Now Laplace's correction factor turns out to be totally correct (or, at any rate, in total agreement with modern physics). So no contradictions arose for Carnot when he made use of either Laplace's or Poisson's results.

Finally, Clément and Desormes had determined γ experimentally. Their value, precision apart, does not conflict with modern determinations.

Thus, for all results connected with the adiabatic heating or cooling of a gas, whether theoretical or experimental, there was no disagreement between the caloric theory and the modern dynamic theory of heat. Therefore Carnot's work was not compromised by his use of any of these nineteenth-century data.

In fact, the only point of disagreement between the old and the new theories is over interpretation. In the modern view, heat is *generated* by the work done on the gas during compressions and *converted* into work during the expansions. In the caloric theory, heat merely changes from latent to 'sensible' (compressions) and from 'sensible' to latent (expansions). In other words, latent heat totally mops up the effects of work.

That heat should be weightless and subtle (invisible) is bad enough, that it should go latent as well is like adding insult to injury. How would such a fudge-factor ever be detected? If the heat could have been tracked *all the way around the cycle*, then differences between the caloric modern theories would have shown up. It is curious that in the *Reflections* Carnot, in contrast to his successor, Clapeyron, never stressed the constancy of this heat around the cycle. Rather, he stressed the identicality of a given point on the cycle, a given *state*, however many times it was reached and from whichever direction. As always, his instincts were faultless.

History tells us (as do Carnot's unseen notes) that the caloric theory would only be ousted by looking at the whole of physics and not just gases. For example, there were problems such as explaining ordinary and radiant heat in one theory; the heats of chemical reaction; and frictional heating, such as occurs in Rumford's experiments. Carnot, in the posthumous notes, suggests repeating Rumford's experiments. Significantly, he also suggests taking the next step that Rumford never contemplated—quantifying the work done by the cannon borer. It is an irony that the arena of gases, so long an aid to the study of heat, was now holding back progress.

This is only one of many ironies. Some others include the following:

- Daniel Bernoulli (in 1738; see Chapter 7, Part III) was the first to consider the scenario of a gas in a cylinder with a piston, and so brought forth the first kinetic theory of gases. Carnot used the same scenario, but assumed an essentially static gas theory (as implicit in the caloric theory).
- The caloric theory both helped and hindered understanding. The model of heat as a fluid helped to make the analogy with water-engines. However, the caloric theory's axiom of heat conservation held back progress.
- The concept of latent heat was the perfect 'fudge-factor'. It managed to almost completely mask the erroneous axiom of the conservation of heat.
- The abstract construct of the ideal heat-engine managed to further the understanding of real, down-to-earth engines.
- Real, down-to-earth engines were busily and noisily converting heat into work all over Europe even while the natural philosophers were debating whether such conversions were possible.
- The ideal heat-engine was reversible, but it was to bring in the Second Law (see below), which would show that all real processes were irreversible.
- While Carnot's syllogism vetoed a perpetually acting machine, Newton's First Law of Motion required perpetual motion. However, the Second Law of Thermodynamics would eventually show that perpetual motion can never occur *in practice*.
- The Second Law would make it harder to oust the caloric theory and to enable the discovery of the First Law of Thermodynamics (the fact that there are always 'losses' in practice means that the 'energy books' don't always appear to balance; also conversions from heat to work occur less readily than from work to heat).

Perhaps, wherever there is a proliferation of ironies, this shows that while some premises are fundamentally right, others are fundamentally wrong.

In assessing the importance of Carnot's work and of physics in general, we note that there has been another less well-known Sadi Carnot—in fact, Carnot's nephew—who was President of France during the Third Republic.

But how can things such as engines and their efficiency, which are so anthropocentric, have anything to do with *physics*? This question will be resolved when we discuss the sequel to Carnot and the birth of thermodynamics in Chapter 16.

Carnot always had an intuition, a hunch, for what was right. But hunches—common sense—so often lead to falsehood. Perhaps the message to be drawn is that one should always follow one's hunches, but only if one is a genius, like Carnot.

13

Hamilton and Green

What I tell you three times is true.
Lewis Carroll, 'The Hunting of the Snark'

The advancing idea of energy required not just the discovery of new 'blocks' but also an increase in the mathematization of the concept. This occurred from the work of two men who were contemporaries but who knew nothing of each other.

George Green

In this chapter, we have yet another case of William Thomson rescuing a posthumous book from obscurity. This was George Green's *An Essay on the Mathematical Analysis of Electricity and Magnetism*.[1]

George Green (1793–1841) is something of an enigma. He was a baker and miller from Nottingham, England, had less than two years of education (he left school aged nine), and produced no scholarly work until he wrote a book in 1828, when he was already 35 years old. He had to publish his book privately, after placing an advertisement in the *Nottingham Review* on 14 December 1827:

> In the Press, and shortly will be published, by subscription, An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism. By George Green. Dedicated (by permission) to his Grace the Duke of Newcastle, K. G. Price to Subscribers, 7s. 6d. The Names of Subscribers will be received at the Booksellers, and at the Library, Bromley House.[2]

He received only 51 subscriptions.[3] The book turned out to be a major work in mathematical physics.

There are no notebooks or manuscripts (and no portrait)—nothing to indicate what Green's early motivations were. Most puzzling of all is how Green gained access to the source material that he quotes, chiefly the *Mécanique Céleste* (*Celestial Mechanics*) of Laplace, and Poisson's papers. Such advanced mathematics (and in French) was not being taught in the outdated curricula at British universities, let alone available in Nottingham bookshops. Records show (and see the advertisement above) that at the age of 30 Green joined a subscription library. It was here, presumably, that he could read abstracts of the French work and then place an order for the originals (there was a trade in such books, interrupted only during the Napoleonic wars in the 1810s).

What Green did was to mathematically define the potential energy of an arbitrary static distribution of electric charges. He started from Laplace's equation (see Chapter 7, Part II), which Laplace had formulated to describe the gravitational attraction at any point outside a gravitating mass, and re-applied it to the case of electrostatic attraction. He then extended it to cases where the test point could be within or on the surface of the region containing the source charges. (An identical treatment was also given for static magnetic attractions.)

The important thing for 'energy' is that the role of forces was lessened while that of 'the potential' was emphasized, becoming more abstract and mathematical. Instead of considering the force between two charged bodies and finding their consequent accelerations, rather, space was hypothetically seeded with test charges and their response to their immediate environment was mapped out—by the 'potential function'. Green thus extended the concept of potential energy from its eighteenth-century beginnings in the work of Clairaut and Laplace (see Chapter 7) and the nineteenth-century contribution of Siméon-Denis Poisson (1781–1840). He was, in fact, the first mathematician to define the potential and express it in functional form, $V(x,y,z)$, where x,y,z are the position coordinates:

> if we consider any material point p, the effect, in a given direction, of all the forces acting upon that point, arising from any system of bodies S under consideration, will be expressed by a partial differential of a certain function of [position] coordinates.... [we] call it the potential function.[4]

Knowing this function ($V(x,y,z)$ or $V(r)$) a test mass, charge, or magnet can 'know how to move' merely from its *local* environment—by determining the partial rate of change of V with x,y,z at the current position of the 'material [test] point'. The test particle doesn't need to know where the sources of force actually are. Indeed, Green also proved 'Green's theorem', which shows that a system of sources within a given volume is entirely equivalent to the *effect* of these sources at a surface enclosing the volume.

One person only responded to the publication of the book in 1828. This was Sir Edward Bromhead, a Lincolnshire landowner and Cambridge graduate in mathematics. He was so impressed that he wrote to Green, offering to help him publish his work in the memoirs of the Royal Societies of London or Edinburgh. Green's diffidence and other preoccupations prevented him replying for almost two years (his father had just died and he had a *de facto* wife and many children, and had to do all his mathematics in the hours 'stolen from my sleep'[5]). However, all turned out well in the end and Bromhead launched Green's new career, helping him to publish his papers and to enrol as an undergraduate at Cambridge as a mature student (Green was already 40 years old).

Green published more papers but hardly ever referred back to his own book and so it was quickly forgotten. He became a fellow of Gonville and Caius College (despite the requirement of celibacy Green still qualified, as, although he now had six children, he had never married)—but tragically he died of influenza the following year (1841).

Four years later (in 1845), Thomson was an undergraduate at Cambridge and came across a reference to Green's *Essay*. He tried but failed to find it in any bookshops or libraries. After his final exams and the day before he was due to leave for Paris, Thomson mentioned it to one of his tutors who promptly gave Thomson his own copy.

Upon his arrival in France, the young Thomson showed the book to various members of the Paris Academy: 'I found Liouville at home and showed him Green's Essay, to which he gave great attention. I did not find Sturm at home but I left a card. Late in the evening, when I was sitting…at our wood fire in 31, Rue Monsieur le Prince, we heard a knock, and Sturm came along our passage panting…he said '*Vous avez un memoire de Green, M Liouville me l'a dit.*' So I handed it to him. He sat down and turned the pages with avidity. He stopped at one place calling out, '*Ah voila mon affaire*'.'[6]

Green's *Essay* caused 'a sensation' and Thomson arranged for it to be reprinted in *Crelle's Journal*. This was a German publication and it still took many years before the work became known in Britain. Today, Green is appreciated as being the founder of the concept of the potential function, $V(r)$, and of potential function theory.

William Hamilton

Born in Dublin, Ireland, William Rowan Hamilton (1805–65) was the only son in a family of five (surviving) children. His name was a permutation of his father's name, Archibald Hamilton, and his godfather's name, Archibald Rowan, the famous Irish patriot. From the age of three he was raised and educated by his uncle, curate of Trim, probably because his father, a Dublin solicitor, had gone bankrupt (the father admitted 'I can manage anything but my own money concerns'.[7])

The uncle quickly recognized Hamilton's precocity in mathematics and languages, and started him on a programme of learning that included Hebrew, Latin, and Greek when he was still an infant. In 1818, Hamilton competed against the famous American Calculating Boy, Zerah Colburn. William lost consistently (an unusual experience for him) but later chatted to Zerah and wrote a commentary on why Zerah's methods worked (Zerah himself had no idea).

As a student at Trinity College, Dublin, Hamilton won every prize and passed every examination *cum laude*. When only 22, he was appointed head of the Dunsink Observatory in Dublin. This was not a particularly appropriate appointment, as his talents were not in practical physics. He made the best of it he could (in the early years, at any rate), helped by his sisters, who also lived at the observatory and took turns with the observations.

Hamilton became close friends with the poets Wordsworth and Coleridge. Poetry and mathematics were the wellsprings of his creativity and he considered them as two equally important aspects of the same underlying reality. (Wordsworth tactfully advised Hamilton to stick to mathematics.) Hamilton was introduced to the idealism of the German philosopher Immanuel Kant (1724–1804) by Coleridge. He even sought Coleridge's approval for his atomistic ideas (Hamilton endorsed abstract point centres of force, *à la* Boscovich; see Chapter 7). Hamilton's idealism also extended to his love affairs. His first love represented the ideal and was a life-long passion even while it remained on an abstract plane (she was forced to marry another).

Apparently, Hamilton was a compulsive jotter and calculator: his children remember him writing notes and computations on anything that was to hand, including his fingernails and the shells of his boiled eggs at breakfast.[8]

Hamilton's Mechanics

Idealist or not, it was Hamilton's predilection for formulating any problem in the most general, abstract, and algebraic terms that shaped his physics. He was like Lagrange in this respect and perhaps this is the reason why he tackled the same problem area (mechanics), carrying on where Lagrange had left off. Hamilton revered Lagrange and referred to him as the 'Shakespeare' of mechanics and to his work as a 'scientific poem'.

Hamilton began by looking at the paths of light rays. He knew of Fermat's 'Principle of Least Time' (from around 1630) for the path of a light ray. In Chapter 7 we saw how, over 100 years later, Maupertuis, Euler, and Lagrange had revived such minimum principles and how this had led to a whole new outlook, different from the search for conserved quantities. Hamilton, as we have noted, was always motivated to do things in the most general way possible and he founded a method that managed to *combine* these two approaches: conservation of quantities and minimization of paths.

Hamilton's point of departure was Fermat's Principle, which determined the path of one given light ray with one given starting angle. He extended this to the case of a whole bundle of light rays, starting from one position (a point source) but with a small spread in the values of the starting angle. By tracking the rays through time and joining together the tips of the rays—the 'arrowheads'—at any given time, Hamilton defined a surface of simultaneous arrival time. He could prove that the rays always pierced this surface at right angles (this was, in fact, a consequence of the Least Time Principle) and that this was true whatever time this 'snapshot' of the arrowheads was taken and whatever system of lenses, mirrors, and so on the rays had encountered.

The surfaces that satisfied this requirement were described by a specific function, the 'characteristic function', and this was Hamilton's big discovery in optics (he was only 18 years old at the time). Notice how a *geometric* idea (rays piercing surfaces at right angles) has arisen and been converted into algebraic form; and also how the emphasis has shifted away from the *minimum* time for the path of *one* ray to surfaces of *constant* time for a *collection* of rays.

In his role as head of the Observatory, Hamilton was tackling the problem of perturbations in the paths of the planets. He considered extending his treatment of light to the motion of particles and devising an 'optico-mechanics' that would be even more general than Lagrange's mechanics. The use of one theory for both optical and mechanical systems was a novel idea although, unknown to Hamilton, there was one remarkable precursor in the work of Johann Bernoulli in 1697.[9] Johann Bernoulli had treated the motion of a particle falling under gravity as analogous to light passing through a medium with a varying refractive index.

In the mechanical as opposed to the optical case, Hamilton replaced the Principle of Least Time by the Principle of Least *Action*. It is noteworthy that he attached no teleological significance to these minimum principles: 'Although the law of least action has thus attained a rank amongst the highest theorems of physics, yet its pretensions to a cosmological necessity, on the ground of economy in the universe, are now generally rejected.'[10]

(In the following, T and V will refer to the kinetic and potential energies, respectively, and not to temperature or volume.)

To start with, Hamilton adopted Lagrange's formulation of the Principle of Least Action whereby the integral $\int L\mathrm{d}t$ was minimized (see Chapter 7, Part IV). Hamilton then replaced L by $(2T - H)$, where the function H, the 'Hamiltonian', denoted the total energy of the mechanical system. This all looks innocent enough until we note that H is the total energy, given by $(T + V)$, and (from Chapter 7) $L = (T - V)$. So Hamilton had merely changed $(T - V)$ to $[(T + T) - (T + V)]$. This has the same whiff of absurdity about it as d'Alembert's rearrangement of $F = ma$ to $F - ma = 0$ (see Chapter 7, Part IV).

To add to the absurdity, we also remember from Chapter 7 that the minimization of $\int L\mathrm{d}t$ is only what *would* have been called Lagrange's Principle if Lagrange had stated it, which he didn't, and is now known as Hamilton's Principle, of which it is just a special case.

The absurdity is removed when we understand that Hamilton and Lagrange weren't considering exactly the same scenario or using the same coordinates or boundary conditions. What Hamilton had done was to *transform* Lagrange's problem into a totally different one. The quotation at the start of the chapter may be aptly paraphrased to: 'What I transform [at least] three times is [still] true.'

The first transformation, actually carried out by Lagrange, was to switch from everyday 3D space to 'generalized' coordinates in 'configuration' space (the terms are explained in the Lagrange section of Chapter 7, Part IV). The second transformation was due to Hamilton and was from configuration space to 'phase' space. The third transformation (or rather, infinite set of transformations) was from one state of phase space to the next, in time.

Let's run through this again more slowly. Lagrange had the degrees of freedom of the given mechanical problem represented by generalized position coordinates, q. He allocated appropriate starting values to the q and also to the corresponding generalized speeds, $\dot{q}$ (a dot over a symbol will indicate differentiation with respect to time). Then applying the Principle of Least Action led to Lagrange's equations of motion: a set of second-order partial differential equations, one for each q. Finally, solution of these equations led to a single world-line in configuration space.

Now, Hamilton kept Lagrange's position coordinates, q, but when it came to their associated speeds, $\dot{q}$, he transformed these and re-labelled them p, and called them 'momenta'. In the old system, each q had its corresponding speed, $\dot{q}$, whereas in the new system each q had its corresponding momentum, p. Finally, instead of assigning specific starting values to the momenta, p, he brought them into the problem as extra variables. So, as compared to Lagrange, Hamilton had a mechanical scenario with twice as many variables (*q and p*). This led to twice as many equations, but one signal advantage over Lagrange was that Hamilton's equations were only *first*-order differential equations. Here is another whiff of the absurd: how can changing from $\dot{q}$ to p (like disguising the differential nature of the ordinary speed by calling it v) really make any difference?

To explain this, we must appreciate that the 'space' of (p,q) values (Gibbs was later to call it the 'phase space'; see Chapter 17) is a different sort of beast to the configuration space of Lagrange. The (p,q) pairs span the entire range of *possibilities*, whereas the configuration space identifies one *actuality*.

A fair analogy is the case of two golfers at the driving range.[11] The first golfer takes a shot, and if we know the starting conditions (the speed and angle of the club, the position of the tee, etc.), then we can solve Lagrange's equations and find the world-line in configuration space and then convert this to an actual trajectory in ordinary 3D space. The specific dynamical problem has thus been solved. The second golfer is a robotic golfing machine. Now the robot doesn't get bored or tired and it can be set to hit 250 balls, one after another. However, as with any mechanism, the exact same starting conditions can't be replicated every time. There will be a small spread in the initial speeds, positions, and angles. We can come back later and see where all the golf balls have landed. We no longer predict the trajectory of one specific ball but chart the range of possible outcomes. This still provides useful information of a different, more general, kind. (For example, the managers of the golf club might note that for the seventeenth hole all the balls are landing in the rough.)

By plotting all the possible (p,q) values as they change in time, a very suggestive metaphor appears—that of a fluid flowing in phase space. This is indeed a subtle sort of fluid as its streamlines don't correspond to real trajectories. However, the multiplicity of the streamlines meant that Hamilton could recast the problem in the same way that he had done for light. Instead of trying to determine one actual trajectory using a minimum principle, he investigated a bundle of streamlines and looked at the surfaces of constant action. He again (as in his optics) found a characteristic function linking successive surfaces. In fact, for infinitesimally small changes, Hamilton's special function could be used to *generate* the next surface. The whole problem of motion was thus *transformed into one of geometry*: time passing was like a succession of infinitesimal coordinate transformations, mapping the phase space on to itself.

The metaphor of a fluid was of enormous utility. It brought out various conservation theorems, such as Liouville's Theorem (for the conservation of the volume of an incompressible fluid) and Helmholtz's circulation theorem (conservation of vorticity). What of the conservation of energy? What of energy itself?

Hamilton's generating function, connecting surfaces of constant action, was intimately connected to the total energy, the 'Hamiltonian' function, H. This is hardly surprising, given that action itself is 'energy × time'. But Hamilton succeeded in generalizing and extending the concept of energy. First of all, it didn't need to be restricted to the form $(T + V)$ (most generally, it is $H = \sum(p\dot{q}) - L$); secondly, it didn't need to be conserved. All that mattered was that between successive surfaces of the fluid in phase space the increment in the action was a constant.

In Lagrange's mechanics, $(T - V)$ seems to loom larger than $(T + V)$. But this is only because Lagrange hid the total energy away as an assumed constraint—it had to be conserved at each and every instant of the motion. Hamilton brought the total energy out into the open. It could be used in more general cases where the kinematical conditions (T) or the potential (V) depended explicitly on the time, and even for cases where the total energy wasn't conserved. This didn't mean that the important law of energy conservation was being abandoned, but that the investigation could include a system that wasn't closed and where energy leaked in or out in a controlled, prescribed fashion (for example, the swinging of an anchor could be described even while the anchor was being slowly wound up). In fact, the total energy within the system, H, was now more than ever the chief determinant of the mechanical problem.

In closed conservative cases (where everything is independent of time), then Hamilton's Minimum Principle does reduce to Lagrange's Minimum Principle. There are two other conditions that must also be satisfied: T must have the usual 'quadratic form' ($T = ½mv^2$) and V must not depend on the velocities. In the more general cases where, say, V does depend on velocity (as turns out to be the case for electric charges moving in a magnetic field, discovered at this time; see Chapter 15) or where V or T does depend on the time (say, the rug is pulled out at constant speed from under my feet)—then Hamilton's mechanics will still be applicable. So, Hamilton's mechanics shows us that energy is too important a concept to be limited to just those conservative or velocity-independent cases.

Contemporary Reception

Hamilton's work was too abstract and too general to be of much use for contemporary physicists—we have met this before in the work of Lagrange and of Sadi Carnot. His work in optics led to one experimental prediction—that light passing through certain crystals should emerge as a cone. This conical refraction was, with some difficulty, detected (by Humphrey Lloyd[12] in 1832) and was received with much acclaim. However, in the main, Hamilton's mechanics paid dividends only in the twentieth century, especially in the area of quantum mechanics (see next section).

The trouble was, the mathematics was difficult and also Hamilton was not a good expositor of his own work. For example, he sent a 20-page letter to the astronomer John Herschel at the Cape of Good Hope (Herschel was observing the Southern skies) giving an enthusiastic account of his method. Herschel replied: 'Alas! I grieve to say that it is only the general scope of the method which stands dimly shadowed out to my mind amid the gleaming and dazzling lustre of the symbolic expressions... I could only look on as a bystander, and mix his plaudits with the smoking of your chariot wheels, and the dust of your triumph.'[13]

However, there was one contemporary mathematician who fully understood Hamilton's mechanical theory and appreciated its amazing generality and beauty. This was the German mathematician Carl Jacobi (1804–51). In fact, Jacobi appreciated Hamilton's work so much that he referred to Hamilton's equations as the canonical equations and the p and q variables as the canonical coordinates. ('Canonical' meant having the same authority and setting the rule in the same way as canon law governed Church procedure—a curious choice of adjective for one of Jewish background). Later, Thomson and Tait complained that they didn't know what was canonical about it.[14]

Jacobi turned Hamilton's mechanics into something of practical utility—he found a way of solving the canonical equations. (Hamilton's theory, while beautiful and abstract, was mostly insoluble as it stood.) Jacobi, at the expense of some beauty, managed to reduce it to one grand equation, now known as the Hamilton–Jacobi equation.

Future Developments

Electromagnetism

Green's potential function, $V = V(r)$, was a more sophisticated mathematical object than, say, $V = mg\Delta h$ (the change in gravitational potential energy between heights zero and Δh). The force on a test mass or electric charge could be determined at a *point* knowing only the *local* conditions (the partial rate of change of V with position at that point) rather than having to refer to a whole distance, such as an interval in height, Δh.

Green had dealt with a static distribution of electric charges. Contemporaneous with his mathematical work, the experimentalists were uncovering a whole new arena—magnetic effects arising from moving charges (currents). The first such effect was Oersted's observation in 1820 of a compass needle moving in response to a nearby current of electricity. Almost immediately, this was further investigated by Ampère in France and Faraday (see Chapter 15) in England. The amazing findings were that the 'induced magnetism' depended on the speed and direction of the charges and the resulting magnetic force was in yet another direction, at *right angles*—all very non-Newtonian results. Faraday, in the 1830s, showed how the magnetism varied from point to point in space and could be represented by 'field lines', which showed how tiny bar magnets (iron filings) would align themselves. A similar 'field' description applied to electricity.

In this new description, each point in space would now have to be flagged with direction and speed, as well as with some function of the position coordinates. But Green's potential function theory could be easily adapted to meet the needs of the new *field* theory.

Between 1861 and 1865, James Clerk Maxwell developed his field theory of electromagnetism (see Chapter 17). He showed that there is energy in the field and it can be transported by electromagnetic waves, of which light is an example.

Light

As we have seen, Hamilton's theory of mechanics started from optics. He claimed that his optics was neutral with regard to the nature of light—in other words, whether light was a wave or a particle. While the French, led by Fresnel, had already come round to the wave theory of light over a decade earlier, the British were still discussing this at British Association meetings such as the one in Manchester in 1842. (Hamilton was one of the founders of the British Association for the Advancement in Science, which Charles Dickens satirized as the association for the advancement of 'umbugology and ditchwateristics'.[15]) The debate in Manchester was getting heated and Hamilton, trying to bring the temperature down, risked a joke, saying he 'hoped it would not be supposed that the wave men were wavering, or that the undulatory theory was at all undulatory in their minds'.[16]

Privately, Hamilton supported the wave theory of light. The seventeenth-century natural philosopher Huygens (see Chapter 3) had put forward a wave theory in which light rays emanated from a point source in all directions and the consequent 'wavefront' was the surface of simultaneous arrival time of all the rays. Also, each point on the wavefront was a new point-source of rays. We can see that Huygens' theory was eminently suited to Hamilton's optics.

Quantum Mechanics

If Hamilton's optics had brought out the wave nature of light, then would his mechanics bring out the wave nature of particles? This provocative question was not even asked until the beginning of the twentieth century. First came the various landmark investigations of Max Planck, Arthur Compton, and Albert Einstein, which showed that while light was wave-like, it also came in particles ('quanta'). The reverse case, that particles could be wave-like, was finally postulated by Louis de Broglie in 1924 (and experimentally shown in the case of electrons by Clinton Davisson and Lester Germer in 1927).

De Broglie cited the minimum principles of Maupertuis and Euler as his inspiration. Only one mathematician was actively promoting Hamilton's great optico-mechanical synthesis and that was Felix Klein (in the 1890s). Klein bemoaned the fact that his assessment of Hamilton had lain unread in the reading room at Göttingen for 20 years and that Jacobi's work had 'snatched away'[17] the glory from Hamilton. However, it only takes one person at a time to keep a thread going and eventually Erwin Schrödinger, through Arnold Sommerfeld (physicist and clerk of the Göttingen reading room), came across Klein's work on Hamilton.

Finally, in 1926, Schrödinger formulated his famous wave equation for a system of masses of microscopic size (for example, he considered the hydrogen atom) and he specifically acknowledged Hamilton: 'His [Hamilton's] famous analogy between mechanics and optics virtually anticipated wave mechanics. The central conception of all modern theory in physics is the "Hamiltonian"...Thus Hamilton is one of the greatest men of science the world has produced.'[18] The energy function, H, is embedded into the heart of Schrödinger's equation.

Now we must remember that in quantum mechanics we are in a different realm—the particles are miniscule compared to golf balls or planets. 'Miniscule' can be quantified: whenever the lengths and momenta yield a quantity of action of the order of Planck's constant, $h = 6.63 \times 10^{-34}$ J s, then we are in the quantum-mechanical world. In this world, some totally new features emerge—the probability wave and Heisenberg's Uncertainty Principle—and the order of observations makes a difference.

These three features are closely connected to each other and *all have their origins in Hamilton's theory*. In Lagrange's Minimum Principle, one varied path was selected in preference to all the others. Now, in the quantum-mechanical realm, we have a new wave feature—a *multiplicity* of paths 'off the minimum', which also have a probability of being followed. In other words, instead of Hamilton's *possibilities* (for, say, 250 golf balls) we now have *probabilities* for just one quantum-mechanical particle.

This multiplicity of paths for just one particle leads to a smearing in the values of position and momentum and to a constraint on the simultaneous determination of position and momentum for that one particle. This is the famous Heisenberg Uncertainty Principle. What is not generally well known is how this Principle arises from Hamilton's variational mechanics. First, we must be aware of yet another change in metaphor: what were 'displacements' or 'variations' in the classical mechanics have now become 'observations' in the quantum domain. Secondly, whereas in the Principle of Virtual Work (Chapter 7) we only considered 'displacements' in one dimension, in Hamilton's Principle we now have two variables, p and q, and we vary each in turn.

Then, the French mathematical physicist Siméon-Denis Poisson (1781–1840) found something curious; namely that the *order* of variations made a difference. Specifically, his so-called 'bracket relation' had a non-zero value when p and q were for one and the same body. However, something useful could be salvaged—while the Poisson bracket was finite, at least its value was invariant (the same for different coordinate representations). Poisson noticed this, but didn't know what to make of it. In Hamilton's theory, it turns out that the Poisson bracket is an invariant for just those transformations that are canonical—in other words, just the important ones that generate the time-evolution of the system.

In quantum mechanics, the Poisson bracket takes on even more significance. It is more than just a test of the canonicity (!) of the transformations, but is the very cornerstone of the new quantum theory. This link between classical and quantum mechanics was emphasized by the twentieth-century physicist Paul Dirac (1902–84). It indicates that the p and q variables (for any one particle) are intimately connected—they are called 'conjugate variables'—and that one variable can't be pinpointed without sacrificing knowledge of the other. 'Energy' and 'time' (for a given particle) are also conjugate variables—like the p and q, the E and t can be multiplied together to give the 'action'. The Uncertainty Principle therefore also applies to energy. Thus, for Heisenberg's inequality (Equation 17.3 in Chapter 17), the energy of a given particle can't be known to better than $h/(4\pi\Delta t)$, where h is Planck's constant and Δt is the uncertainty in the time.

From this rather technical discussion, we can extract some interesting dividends. We are familiar with the idea that, in quantum mechanics, classical certainty about position and momentum must be sacrificed. But now, going the other way, we find that in the classical mechanics some mathematical formalism (the Poisson bracket) is almost empty of meaning until the quantum realm is reached.

Metric Theory

We have already found some amazing developments stemming from Hamilton's theory, but there is more to come. Instead of solving differential equations of motion, Hamilton's (p,q) space suggested a totally new way of looking at a mechanical problem. We need no longer think of the p and q coordinates as functions of time, but we can consider them as points in phase space and nothing else. We can even make the problem 'static' by treating the time as just another variable (admittedly, this has the feeling of another trick and yet another transformation) and label each point with t as well as with p and q. We then look at the (geo)metric features of the given 'space'; that is, the invariants. (We have already seen that the Poisson bracket is *the* invariant quantity in the (p,q) 'space' of quantum mechanics.)

Or we can turn this around and *start* with the invariant quantity and then investigate which transformations will guarantee this invariance. (This was what Hamilton did when he demanded that the mechanical paths in (p,q) space maintain the 'geometric ray property'.)

In Einstein's Special Theory of Relativity, the speed of light, c, is the new invariant quantity and the coordinate transformations that ensure that c remains invariant are the Lorentz transformations. Speed is still 'distance/time', so if c is the new invariant then this is at the expense of distance intervals and time intervals that are now no longer invariant and also get 'mixed up'. Einstein's Theory also showed that another quantity was invariant (mc^2) and that mass and energy were related by $E = mc^2$ (see Chapters 17 and 18). Thus entered one of the most remarkable discoveries of all physics—the equivalence of mass and energy—and all in order to satisfy the overarching requirement that c is a constant for all observers (frames of reference).

In Einstein's General Theory of Relativity, the agreement between different observers is again paramount, but what counts as a valid observer has been extended from a reference frame with constant speed and direction to one that can be accelerating (when viewed over sufficiently short times and distances). The result of this is a blurring of the distinction between mass and 'geometry'. We have already made reference to this at the end of the section on d'Alembert and Lagrange (see Chapter 7, Part IV) in the slogan 'mass tells "space" how to curve; curved "space" tells mass how to move'. Now that we have an equivalence between mass and energy, there is a further twist—'curvature in space' is a *source* of mass-energy and mass-energy is a *source* of 'curvature'.

Overview

We have heard that energy is in the electromagnetic field, energy is quantized, energy is 'uncertain', energy and mass are equivalent, and energy and curvature are sources of each other. Of course, Albert Einstein (1879–1955) and certain predecessors and contemporaries must take the credit for these remarkable ideas. However, both Green and Hamilton paved a bit of the way by advancing the mathematization of concepts and, in particular, moving towards a field theory (Green) and a metric description of reality (Hamilton).

But how can defining new mathematical objects or carrying out lots of coordinate transformations really make any difference? In order to understand this, we must appreciate that these mathematical advances represent more than just techniques and conventions—they correspond to real physical discoveries: energy from the Sun *does* reach us every day and *is* transported across empty space by varying electromagnetic fields; the energy locked away in mass is real and was released with tragic consequences in Japan in 1945. The mathematics shows us a new reality that we wouldn't necessarily uncover just from direct observations.

As regards energy, the legacy of Green and Hamilton was in the new mathematical functions, the potential function, V, and the Hamiltonian, H. Green's work was relevant to that of Hamilton, as having V in functional form allowed H to be in functional form. (It is not known whether either man knew of the other's work.) H and V are still abstract, unspecified functions—the physicist must decide what goes into H or V in any given physical scenario.

In Chapter 7, we contrasted Lagrange's 'systems approach' with Newton's 'trajectory of a particle subject to different forces' approach. Now, with Hamilton, the systems approach is taken one stage further. Lagrange's system represented by a point moving along its world-line in configuration space is replaced by Hamilton's 'system of systems', whereby all possible trajectories in phase space are examined.

Furthermore, with Hamilton, 'total energy' plays an even more important role than it did in Lagrange's mechanics, and has been freed up from being a mere conserved quantity to being the chief determinant of the system. It is not that energy conservation has been sacrificed, but rather that Hamilton's methods allow non-conservative systems to be analysed in cases where the time-dependence is in functional form (for example, the system might be an anchor swinging on its chain, but all the while the chain is being hauled in). In other words, energy is too important a concept to be restricted to just those cases where it is conserved. Also, energy has been generalized from the usual ($T + V$) to a quantity where T doesn't need to be in 'quadratic' form (i.e. depend on the square of the speed) and V can also depend on the velocity.

Hamilton's variational mechanics barely caused a ripple in his own era, but came into its own in the case of quantum mechanics applications. The problems here are almost always too difficult to solve in any particular case and the phase space mapping of possibilities is therefore especially useful in delimiting the problem. Coordinate systems can be specially selected to bring out any inherent symmetries or conserved properties, making the problems tractable or, at least, more physically meaningful. For the same reasons, Hamilton's methods have also been useful in statistical mechanics. Here, there may be a system with 10^{30} particles or more (say, a mixture of gases), so doubling the number of equations to solve in going from Lagrange's equations to Hamilton's equations is no big disadvantage.

'Action' is always the important quantity, but, after Hamilton, it can be sliced up as (p,q) as well as by (E,t). In the (p,q) slicing, both the p and the q have some 'extensive'* aspects; that is, both are both 'fishy' and 'fowly'. On the other hand, in the (E,t) slicing it seems as if E has pared off all the extensive aspects, leaving naked t all to itself.

* Extensive aspects are those which alter as the sample size is altered (e.g. mass, length, and volume); intensive aspects remain unaltered (e.g. density and temperature). I am not sure whether 'time' is to be considered as extensive or not.

14

The Mechanical Equivalent of Heat

It is the 1840s and the time is at last ripe for the discovery of energy. We have heard of men such as Daniel Bernoulli, Sadi Carnot, and William Rowan Hamilton, who were years ahead of their time. But now there is a growing appreciation of energy, more often called 'force' ('Kraft' in German), its interconversion and its conservation. Men such as Ludwig Colding (in Denmark), Julius Robert Mayer and Hermann von Helmholtz (Germany), Marc Seguin (France), and James Joule (England) were independently arriving at the same discoveries. We shall only cover the work of the two key players, Mayer and Joule.

The time was ripe for discovery, yes, but not for a ready acceptance of the new ideas. Both Mayer and Joule started out on the fringes of the scientific establishment and it was many years before their work was appreciated. The trouble was, scientists of the day still found it hard to accept a quantitative link between heat and mechanical energy. It was, by now, generally agreed that heat was a sort of motion and that heat was a by-product of collisions, friction, chemical reactions, and so on. But these were all qualitative results. That there could be a *quantitative* link was hard to contemplate. It was like a category error—like comparing, say, p.s.i. (pounds per square inch) and psa (pleasant Sunday afternoon).

Julius Robert Mayer

Many of the new thinkers were coming from Germany. Whether a trend towards the unification of Germany (from 1848 onwards) also ushered in a unification of ideas is an interesting but unanswerable question. What is certain, however, is that there was a German way of thinking even before there was a country called Germany. Specifically, the maxim 'cause equals effect' was invoked and given central prominence in the work of three Germans in energy physics: Gottfried Leibniz (Chapter 3), Mayer (this chapter), and Helmholtz (Chapter 15).

Julius Robert Mayer (1814–78) was the son of an apothecary in Heilbronn, South Germany. He studied medicine at Tübingen and was an average student, but with a fiercely independent spirit (he belonged to a forbidden, secret society and when he was banned from studying for a year he protested by going on a six-day hunger strike). He eventually qualified as a doctor and, against parental advice, took employment as a ship's doctor on the Dutch vessel *Java*, bound for the East Indies.

The ship set sail on 22 February 1840, from Rotterdam, and the voyage took three months. There wasn't much to do on board and Mayer records in his diary that he had little association with the ship's officers and spent much time reading his science books and feeling hungry. Upon arrival at Surabaya, there was at last something medical to be done and this provoked the incident that was epiphanous for Mayer. He had to let the blood of some sick sailors and he noticed that their venous blood was uncommonly bright red—more like arterial blood. He was informed that it was always like this with new arrivals in the Tropics but, in a flash of insight, Mayer suddenly saw the whole picture: the redness of the blood was due to the balancing of 'force' in natural processes. The air temperature was high in the Tropics and the body therefore had less need to deoxygenate the blood in order to maintain body temperature.

Mayer quickly saw the generality of this idea—the conservation of 'energy'—and applied it to as many physical processes as he could think of. For example, he learned from a local navigator that the sea was warmer after a storm—so, evidently, the motion of the waves was converted into an equivalent amount of heat…

Mayer went into a meditative state, barely exploiting his chances for shore-leave (this was a source of jokes amongst the crew) and took up the train of thought that was to dominate his scientific career—in fact, the rest of his life.

Upon his return to Heilbronn, Mayer acquired a large medical practice, was appointed town surgeon, and subsequently married and had seven children (five died in infancy); but he was a man with a mission—he wanted to understand and promote his new idea of 'force' (energy).

Mayer was a philosopher rather than an experimentalist and his newly conceived philosophy was: nothing comes from nothing; cause equals effect; whenever a 'force' is consumed, the same amount of 'force', possibly in another guise, is generated. He gave as his prime example the conversion of 'fall force' (gravitational potential energy) into 'moving force' (kinetic energy) and quickly generalized this to all other processes in physics: 'motion, heat, light, electricity and the various chemical reactions are all one and the same object under differently appearing forms'.[1]

Mayer observed that total mass was always conserved in chemical reactions, whatever mutations had occurred, and this could be demonstrated by careful weight measurements: 'stoichiometry falls into our lap like ripe fruit'.[2] But Mayer also observed that 'force' was to physics what mass was to chemistry. Surely, then, careful *quantitative* determinations of 'force' would make it reveal itself.

While aware of the many possible transformations, it was in particular the transformation of heat into motion that spoke to Mayer. Motion could disappear (as in an inelastic collision) but total 'force' could never be reduced to nothing—therefore the invisible 'force of heat' must have been generated and must exactly (quantitatively) make up for the loss of motion. Mayer calculated that a one-degree rise in temperature of 1 kg of water represented exactly the same amount of 'force' (energy) as the 'fall-force' in a 1 kg mass released from a height of 365 m (or 1,000 Paris feet) (see later). Motion had disappeared and heat had been generated, but there was no need to understand how this transformation occurred or what heat actually was at a more fundamental level—its calorimetric measure was sufficient. The approach is perfectly captured in Feynman's analogy of Dennis' blocks—the formula's the thing.

While the fundamental nature of heat didn't enter into the calculations, it wasn't likely that it could be material as heat had to appear and disappear at a stroke (in strict subservience to the conservation of total 'force'). Mayer was the first to expose the nakedness of the Emperor when he proclaimed: 'the truth—there are no immaterial materials'.[3]

Mayer wrote four papers in quick succession. The first, in 1841, was straight after his return from the East Indies. It was rejected by Poggendorf and the original manuscript was not even returned. This was disappointing, but gave Mayer time to improve his shaky knowledge of mathematics and physics. The second paper (in 1842) was much better, included the mechanical equivalent of heat (see below), and was published by Liebig—but in the section on chemistry and pharmacy, so it sank almost without trace. The third, in 1845, was rejected by Liebig, and Mayer had it printed privately, at his own expense, by the bookstore in Heilbronn.

This third paper was almost book-length (112 pages) and was a highly original exposition of energy transformations in living processes. Contemporary physiologists, such as Liebig, did by now (1840s) understand that animal heat arose from the combustion of food (as opposed to earlier ideas, such as heat being due to the friction of the circulating blood). However, Mayer was the first to consider the totality of energy transformations—not just body heat, but the work done by the animal, heat losses due to friction, and so on. In other words, the calorific value of the food had to account for *all* the energy conversions and not just for maintaining body temperature. Mayer also considered the cosmic role of the Sun and the energy transformations occurring in (what we now call) photosynthesis, transpiration, and so on. Sadly, the title of the paper, 'The Motions of Organisms and their Relation to Metabolism',[4] didn't convey its contents and the paper was largely ignored.

Depressed but undaunted, Mayer printed a fourth paper, again at his own expense, in 1848. This one, on celestial dynamics, was also highly original, and put forward hypotheses for the source of the Sun's energy (meteors falling in); the bright tails of shooting stars (friction in the atmosphere); the effect of the tides in slowing down the rate of rotation of the Earth; the increasing rate of the Earth's rotation as its volume was reduced by cooling; and other ideas.

But 1848 was a bad year for Mayer. It was a year of revolution in Europe and Mayer's conservative attitudes led to his being briefly arrested by insurgents and permanently estranged from his more rebellious brother, Fritz. Two of Mayer's children died and also a crank named Seyffer ('nothing'?) ridiculed Mayer's heat-to-motion conversion in the newspaper. After years of almost total neglect, and the fact that others were now beginning to take the credit for similar ideas (principally Joule, but also Liebig, Holtzmann, and Helmholtz), this was perhaps the last straw for Mayer. In May 1850 he attempted suicide, jumping from a third-floor window. His feet were badly damaged but eventually recovered. His mental state wasn't so easy to fix and Mayer voluntarily admitted himself to a private sanatorium. Unfortunately, this led to a loss of autonomy and to a number of forced admissions to various mental institutions where Mayer was treated rather badly (e.g. he was made to wear a strait-jacket). He eventually made a complete recovery, but was out of the scientific scene for almost a decade.

The story has a moderately happy ending. Mayer re-emerged in around 1860, a year that coincided with the time when his work finally began to be recognized. Helmholtz and Clausius had discovered Mayer's papers and lauded him as the true founder of the energy principle. Through Clausius, the English physicist John Tyndall came to hear of Mayer and championed his cause against the chauvinistic claims of Peter

Guthrie Tait (on behalf of Joule). William Thomson stayed in the background, but his sympathies lay with Joule. Between Mayer and Joule themselves there was no animosity: Mayer was full of admiration for Joule, but Joule was more guarded in his appreciation of Mayer (see the end of the next section).

Mayer's enduring legacy was his vision, in Java, of a single, fixed quantity of indestructible 'force' in nature and that heat and motion were but manifestations of it. He carried out no experiments but correctly identified the difference between the specific *heats* of a gas, $C_P - C_V$, as a quantitative measure of the (external) *work* done when a gas expands adiabatically. By using pre-existing data (that of Delaroche and Bérard, and Dulong; see Chapter 10), Mayer was able to determine the first value for the conversion between heat and mechanical work, the so-called 'mechanical equivalent of heat'. His value[5] was 365 kg-m kcal^{-1}. That is, the 'force' in a mass of 1 kg falling from a height of 365 m was equal to the heat required to raise the temperature of 1 kg of water by 1°C. (This corresponds to 3,580 J kcal^{-1} and compares to the modern value of 4,186 J kcal^{-1}.) Mayer was even able to respond to a criticism of Joule's: how could Mayer justify his assumption that *all* the heat had been converted into work—hadn't some of it gone just to cause the expansion of the gas? Mayer replied that he knew of Gay-Lussac's twin flasks experiments (see Chapter 10, Part II), which showed that no heat is consumed when a gas expands freely, or in other words, merely changes its volume.

All this is impressive: Mayer had conquered the 'category error', considered a variety of scenarios (free fall, expansion of a gas, shaking of water, physiology, and the solar system), and realized that quantification was crucial.

Nevertheless, looking at Mayer's original papers one finds parts that are almost incomprehensible to the modern reader. In 1841 Mayer writes: 'The falling of a weight is a real decrease in the volume of the earth and must therefore stand in some relation to the heat produced.'[6] In the 1842 paper this is explained: 'If we assume that the whole earth's crust could be raised on suitably placed pillars around its surface, the raising of this immeasurable load would require the transformation of an enormous amount of heat…But whatever holds for the earth's crust as a whole must also apply to every fraction thereof…[therefore] by the falling of [even a small] weight to the earth's surface, the same quantity of heat must be set free."[7] In other words, Mayer likened the heat produced by a falling weight to the heat from adiabatic compression of a gas…

This example jolts one into an appreciation of the difficulties in applying a new philosophy.

James Prescott Joule

James Prescott Joule (1818–89) was born in Salford, near Manchester, into a family of successful brewers. Like Mayer, Joule was outside the scientific profession and, like Mayer, it took many years before his work was noticed, let alone accepted. But in other respects, Joule was altogether more fortunate than Mayer. Joule's father employed a private tutor (for James and his elder brother Benjamin)—not any old tutor but the illustrious John Dalton, founder of the Atomic Theory (Chapter 10). Dalton was not only a famous scientist, but also a born teacher. Joule's father also equipped James with his own laboratory in their house in Salford, and freed him from the obligation to earn a living, although he did have some duties regarding the brewery.

Initially, Joule was so far removed from any ideas about a finite totality to Nature's resources that he was being pulled in the opposite direction, researching an 'electro-magnetic machine' that could out-perform the steam engine and possibly even yield perpetual motion. But how could the genie of perpetual motion surface again when it had been dismissed at the end of the eighteenth century?

Moritz Jacobi (brother of Carl Jacobi; see Chapter 13) was to blame. He proclaimed (in 1835) that electromagnetic engines could provide a source of power that was very likely unlimited—the rotating electromagnets should keep on accelerating as their magnetic poles were attracted to the fixed poles on approach and then repelled as they receded. Perpetual mechanical engines and heat-engines were clearly seen as a no-no; when the wind or water stopped flowing or the coal stopped burning, the engine came to a halt. However, the case wasn't so obvious for the new electromagnetic engine. The battery components got used up, certainly, but this was gradual and somewhat mysterious—it wasn't clear what exactly was powering the engine.

In synopsis, the history of the electric motor is as follows. In 1820 the Danish physicist Oersted had made a serendipitous discovery—'galvanic electricity' in a wire could cause the needle of a magnetic compass to be deflected. This link between electricity and magnetism was immediately explored by the French school (Jean-Baptiste Biot, Félix Savart, François Arago, and especially André-Marie Ampère) and also by Davy's young assistant at the Royal Institution, Faraday. In the 1820s Arago discovered the solenoid, Sturgeon the electromagnet, and Faraday the phenomenon of electromagnetic induction. These discoveries opened up the possibilities of electromagnetic machines to generate electricity (the dynamo) or to generate motion (the electric motor). The prototypes were made by Hippolyte Pixii in 1830 and Salvatore dal Negro in 1831.

The electromagnetic machine had thus been proposed by professional scientists—but thereafter its perfection was entirely in the hands of amateurs. Men from all walks of life—physicians, priests, surgeons, lawyers, teachers, bankers, and one brewer—enthusiastically tried to match Jacobi's promise and in the 1830s an 'electric euphoria' swept across Europe and the United States.

So it was that Joule, a teenager, became enthused. He started by giving himself and his friends electric shocks and by subjecting the servant girl to a steadily increasing voltage until she became unconscious (at which point the experiment was stopped). He took up the challenge of perfecting the 'electro-magnetic machine' and submitted his first paper on this, a letter to Sturgeon's *Annals of Electricity*, in 1837, at the age of 19.

Joule was encouraged by the fact that the 'power of the engine' was proportional to the square of the current (I^2), while the zinc consumption in the battery was proportional only to I. Thus 'the cost of working the engine may be reduced *ad infinitum*'.[8] (We shall use the modern symbols V, I, and R to denote voltage, current, and resistance, respectively, although no standards or units had been developed at this time.) Further research showed that, in fact, the duty *decreased* as the current was increased and Joule's hopes of perpetual electric power were dashed. However, he recognized that electricity might be a useful alternative to steam in special cases—it was cleaner, safer and could easily provide rotative power—and he continued his investigations.

Joule quickly established the first of two relationships—what we now call the laws of electrical *energy*—that the 'power of the engine' (the strength of the magnetic attraction) was proportional to both the number of batteries and the strength of the current. (In modern notation, the power, W, is the product, VI.) He then identified heating in the coils as a waste of power (this wasn't particularly surprising as, after all, heating was also a loss in mechanical engines)—and carried out a systematic examination of the heating in all kinds of voltaic circuits (circuits with either batteries or electrolytic cells). This eventually led to his second electrical energy law: the amount of heat lost in an electric circuit was proportional to I^2R.

This is easy to state but conceals an enormous amount of physics—it wasn't as if a package of 'I^2R' lay waiting to be recognized. In condensed form, what Joule did was as follows:

- Establish standards for a quantity of static electricity, current electricity, resistance, and voltage ('electromotive force' or emf).
- Devise measuring instruments (the galvanometer and voltmeter).
- For metallic conductors, consider the type of metal (copper, iron, or mercury), the length and thickness of wire, and the shape of the circuit.
- For the battery itself, determine the resistance of the battery, the heat lost to the surroundings, the specific heat capacities of the various liquid and solid battery components, and estimate the heat due to the solution of zinc oxide in sulphuric acid.
- For electrolytic cells, estimate, again, the heats of solution but also the heats of dissociation and of gas formation. He also estimated the 'back emf' and how this depended on the material of the electrode.

All this work culminated in Joule's understanding[9] that for any voltaic circuit the heating was proportional to I^2R. Now Joule knew of Ohm's Law (although maybe not with that attribution), that $V = IR$, and so recognized I^2R as equal to VI. He had therefore shown, in two separate series of researches, that the mechanical power of an electric machine was proportional to VI and the heating effect was also proportional to VI. In other words, *the mechanical and heating powers were proportional to each other*.

We are on the very brink of Joule's discovery of the interconvertibility between heat and mechanical power. But Joule went cautiously, step by step. He remarked: 'Electricity may be regarded as a grand agent for carrying, arranging and converting chemical heat.'[10] But how much of the heat was, perhaps, merely transported from the battery and how much was truly *generated* from work? Joule knew exactly what experiments to carry out to clinch the matter. He would investigate electric currents arising purely from mechanical work in a 'magneto', or in other words, with no *chemical* sources such as batteries or cells.

These experiments, carried out in 1843, ended with Joule's landmark paper: 'On the calorific effects of magneto-electricity and on the mechanical value of heat'.[11] Once again, extraneous sources of heating and cooling had to be accounted for: heating due to eddy currents in the metallic cores of the electromagnets; cooling by virtue of the spinning movement of the armature, and so on. When these effects had been separately measured and corrected for, the I^2R law could shine through. There was no chemical source of heat and *no cooling elsewhere* in the circuit. Moreover, when a battery *was* included, the induced magneto currents could be made to enhance, cancel out, or even reverse the battery current—and the resultant heating was always proportional to the square of the net current. All this argued against transference of a material heat-fluid, as how could a fluid be cancelled out? Joule summed it up by saying that in magneto electricity we have 'an agent capable by simple means of destroying or generating heat'.[12]

There was still the question of the work done and whether this maintained a 'constant ratio' to the heat generated or destroyed. To determine this, the 'mechanical force' employed in turning the armature (it was rotated by cords attached to weights via a pulley) was measured. First, the weights required just to overcome friction and air resistance were found by running everything with no current through the electromagnets. Then, with the currents switched back on, the heat generated was measured and all the corrections made, even that due to heat lost because of sparks at the commutators (it is not known how Joule accounted for this).

Finally, as well as measuring the heat generated from work put in (running the electromagnetic machine as a magneto), the machine was run in reverse, as a motor, and Joule determined the work put out from 'heat consumed' (see the discussion at the end of the Joule section). From these experiments, he found a measure for the work done in order to generate the heat that could raise the temperature of 1 lb of water by 1°F—the 'mechanical equivalent of heat'. He found values varying between 587 and 1,040 ft-lb. Averaging these, he concluded that:

> The quantity of heat capable of raising the temperature of a pound of water by one degree of Fahrenheit's scale is equal to, and may be converted into, a mechanical force capable of raising 838 lb to the perpendicular height of 1 ft.[13]

Joule then went on to determine a mechanical equivalent of heat, equal to 770 ft-lb, but from a totally different scenario: mechanical work was done to force water through fine, capillary tubes and the consequent rise in temperature was measured. This was just a 'look-see' experiment.

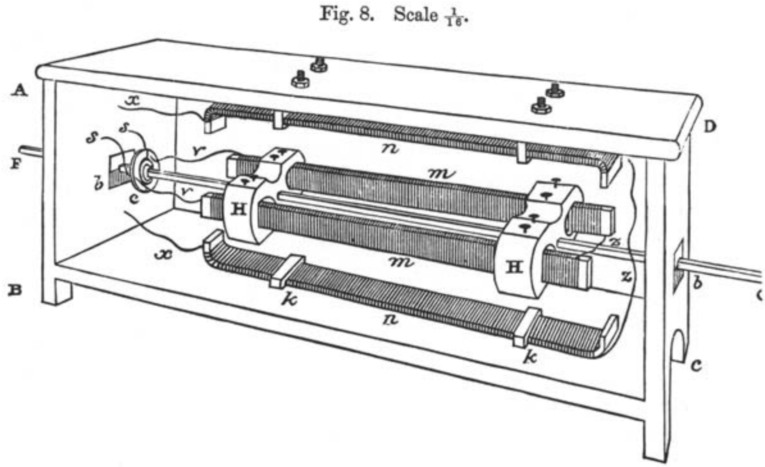

Fig. 14.1 Joule's electro-magnetic engine, from *The Scientific Papers of James Prescott Joule*, volume 1 (reproduced with permission of the Institute of Physics).

However, it was strongly confirmatory of the earlier results with the electromagnetic machine. Joule had no doubts:[14]

> the grand agents of nature are, by the Creator's fiat, indestructible; . . . wherever mechanical force is expended, an exact equivalent of heat is always obtained.

Joule presented these results at the British Association (BA) meeting in Cork, Ireland, in 1843. Very little notice was taken of them. Joule was disappointed but not discouraged—he knew he had uncovered a new principle, and one which would, surely, eventually command the attention of science.

Joule then looked for a different arena in which to test his principle—the compression and expansion of gases (he knew nothing of Mayer's work at this stage). He monitored the temperature changes as air was compressed to or expanded from a very high pressure (22 atmospheres, or 'atm'). There were large errors due to the high take-up of heat by the copper vessels and the water-bath, but Joule also wondered about an error of a completely different sort: was the heat change due entirely to the work done on the piston, or did the gas consume heat just by virtue of its volume change?

Joule tested this by repeating (unknowingly) Gay-Lussac's experiment of 40 years before (Chapter 10, Part II, 'Free Expansions'). He employed twin copper vessels linked by a special stopcock. Air at 22 atm in one vessel expanded into the near-vacuum in the other vessel and no temperature changes were observed (once equilibrium conditions had been reached). From this, Joule drew the important conclusion that a volume change, *per se*, caused no heat change. (Later, this was understood to be strictly true only for ideal gases.) The compression and expansion of a gas, while doing external work on a piston, *could* therefore yield a value of the mechanical equivalent of heat. The values so obtained were consistent with the ones determined already from the electric motor and capillary experiments—but the Royal Society, as before, declined to publish Joule's work.

Joule continued with his now-famous paddle-wheel experiments (1844–5), in which falling weights drive a paddle-wheel that causes water to heat up.

As always, there were experimental challenges. The heating effect was tiny: Joule used his own very high-precision thermometers and could measure temperature changes as small as 1/200°F. Baffles were used to break up the motion of the water and prevent its rotation *en masse*. It

Fig. 14.2 Joule's paddle-wheel calorimeter, 1845 (Science Museum, London/ SSPL).

was important to minimize and identify heat losses to the surrounding air and heat gains from friction at the bearings. When all these various effects had been accounted for, the mechanical equivalent of heat was found to be 890 ft-lb.

Joule presented these findings to the BA meeting in Cambridge in 1845. He suggested that those who 'reside amid the romantic scenery of Wales or Scotland'[15] might like to measure the temperature difference between the top and bottom of a waterfall. He also sent letters to two leading men of English science, Michael Faraday and John Herschel. (In all of Joule's career, there was surprisingly little communication between him and Faraday.) However, Faraday was not in sympathy with such quantitative determinations (see Chapter 15), while Herschel was not a fan of the dynamical theory of heat and wrote (in 1846): 'I confess I entertain little hope of the success of any simply mechanical explanation of the phenomena of heat…you will excuse me if I say that I have no time for the subject.'[16]

By 1847, after no response from his talks, papers, and correspondence, Joule, presumably in some desperation, gave a public lecture at a little church in his home territory of Manchester. He explained that 'living force' and 'attraction through space' were interchangeable, as when fingers wind up a watch-spring. Also, shooting stars glowed because they moved very fast through the atmosphere and their living force was then converted into heat of sufficient intensity to make them burn away.

The lecture was covered by the *Manchester Guardian* newspaper, competing with news items such as the 'shocking murder' at the 'rural village of Chorlton-cum-Hardy, a sweet, quiet spot'.[17] The lecture was well appreciated, but Joule admitted that people who were not 'scientific folk' found it very hard to accept that water could be heated by being agitated; also, his theory of shooting stars was contrary to common experience—objects are usually *cooled* when travelling through cold air.

However, 1847 was to be a good year for Joule. It was the year when he got married and also when he presented his work at the BA meeting in Oxford. Amongst many well-known figures of science attending—George Airy (Astronomer-Royal), John Herschel, William Hamilton, Urbain Le Verrier, John Couch Adams, the Reverend Baden Powell, Charles Wheatstone, George Stokes, and other notables—there was one very junior physicist, William Thomson. It was Thomson who took an interest in Joule's work and, at last, brought it to the attention of science.

(We shall cover the work of Thomson, later known as Lord Kelvin, in Chapter 16.)

It was still almost three years before Thomson came round to accepting Joule's findings, but straight after the meeting Thomson wrote to his brother: 'Joule is, I am sure, wrong in many of his ideas.' Thomson's objections were that Joule's work clashed with Sadi Carnot's theory (which required heat to be conserved), and that Joule had shown the conversion of work into heat but not the conversion of heat into work. These objections were finally resolved by the famous paper of Clausius in 1850 and the subsequent researches of both Thomson and Clausius (see Chapter 16).

Joule and Thomson later went on to form a famous collaboration and discovered Joule–Thomson cooling and the Joule–Kelvin effect, amongst other things (Chapter 10, Part II). Thomson was also the source of a famous anecdote about Joule. Soon after the BA meeting Thomson was walking in the Chamonix district in Switzerland when he had a chance encounter with Joule and his wife on their honeymoon. Joule apparently had with him a long thermometer and was measuring the temperature of a waterfall while his wife was in a carriage, coming up the hill. The encounter really did take place, but the bit about the thermometer is too good to be true—Thomson knew how to improve a story.

Joule had shown that heat and work are interconvertible and that the 'exchange rate' for the conversion is 772.24 ft-lb for a temperature change of 1°F in 1 lb of water (this is the same as 4,155 J $kcal^{-1}$, only 0.75% lower than the modern value of 4,186 J $kcal^{-1}$). Significantly, he had shown this in many different domains (weights falling, electrical heating, the motor and the magneto, water forced through capillaries, expanding gases, and so on) and his work would soon open the door to a new abstract quantity—energy.

It wasn't quite fair for Thomson to complain that Joule hadn't shown the conversion of heat into work—he had demonstrated it in two ways: the cooling of a gas expanding against a resisting pressure and the generation of motion in an electromagnetic machine. However, it *is* true that the conversion in this direction was less evident and somewhat counterintuitive. Take the case of the electric motor. Joule had compared the running of the motor first with it stalled by the addition of weights and then running faster and faster as the weights were progressively removed. He noted that the faster the motor spun round, the less current was drawn and the less was the consequent heating. His conclusion?—that the heat deficit had been converted into the work of the motor.

Nowadays, we find Joule's reasoning almost as alien as Mayer's compressive heating from a falling weight. Instead, we argue the case as follows: as the motor is more and more loaded with weights it has to work harder and harder and so draws more and more current from the battery. The heating increases (as I^2R) and eventually the stalled motor will be burnt out. The heating represents a dissipation of energy and in no way goes to increase the work of the motor. It is not heat but, rather, 'electrical *energy*' that has been converted into work. Joule's reasoning and his calculations are all correct provided that we take the term 'heat' as an alias for 'electrical energy'.

Incidentally, the fallacy in Jacobi's argument (for a perpetually acting electric motor) can now be explained in the following way. As the motor spins round, it acts as a magneto and induces currents in itself and in the fixed circuitry. These currents are always in a direction such as to counter the pre-existing magnetic fields and so always reduce the rate of the motor. Jacobi and also Lenz soon discovered these counter-currents and the effect became known as 'Lenz's Law'.

In a historical account such as this, we are following the trail of the winners, but consider, for example, the tale of the unflagging Professor Wartmann. He investigated (in the 1840s and 1850s) the effect on an electric current of high mechanical pressure and of coloured lights: also, the influence of an electric current on the diffraction and polarization of light. He also looked for a difference in the rate of cooling of electrified and non-electrified bodies. All gave null results. (Faraday finally did show a link between electricity and the polarization of light; see Chapter 15.) Part of Joule's skill lay in having a hunch for what phenomena to follow up (or you might conclude that he was just lucky).

John Waterston

An exact contemporary of Joule, John Waterston (1811–83) was an engineer from Edinburgh, and was another scientist in the tradition of great British lone researchers who had uncannily correct intuitions and struggled for recognition by the scientific establishment.

However, Waterston's scientific career was altogether less fortunate than Joule's.

The kinetic theory had a beleaguered start. Daniel Bernoulli, Cavendish, and Herapath were all independent discoverers (see Chapters 7, 8, and 11) and now Waterston joined the ranks of co-discoverers whose work barely saw the light of day.

Waterston had two biographical details in common with Joule and Mayer: his family were involved in the liquor trade (Sandeman's Port) and he took a job with the East India Company. While in Bombay, Waterston wrote a short book—published anonymously in Edinburgh in 1843—in which the basic principles of his kinetic theory were included. He likened the gas atoms to a swarm of gnats in the sunshine.[18] The book drew little attention, perhaps because of its title: *Thoughts on the Mental Functions*.

Still in India, Waterston followed the book up with a more detailed paper, submitted to the Royal Society in 1845. In it, he foreshadowed many aspects of the modern kinetic theory: equal average kinetic energies even for molecules of different masses; a constant ratio of the specific heats, C_P/C_V; and rotational as well as translational modes of motion.

It was evidently too advanced for the referees, one of whom wrote: 'This paper is nothing but nonsense.'[19] To make matters worse, the rules of the Royal Society meant that the manuscript could not be returned. As Waterston had not kept a copy, he could not try to publish it elsewhere. (It's hard for us to imagine today, what with computers, flash drives, and so on, that an author would not have his or her own copy.) A brief abstract was published in 1846.

Waterston's manuscript finally came to light in 1891, eight years after his death. The physicist Lord Rayleigh saw it referred to in a subsequent paper of Waterston's on sound. As Rayleigh was then secretary of the Royal Society, he had no trouble locating the original in the archives. He recognized its great worth and published it in 1892, adding a caution: 'a young author who believes himself capable of great things would usually do well to [first] secure the favourable recognition of the scientific world by work whose scope is limited, and whose value is easily judged, before embarking on greater flights'.[20]

Waterston had gone on to other scientific work (he estimated the size of atoms to be around 10^{-8}cm and the temperature of the Sun's surface as 13 million degrees[21]). However, after a further rejection in 1878, he shunned all scientific societies and contacts. His death was mysterious—he drowned after falling into a canal in Edinburgh, possibly due to a dizzy spell brought on by heat stroke.

Overview

Although it took some time for the British to accept it, there can be no doubt that Mayer got to energy first. He was awarded the Copley medal of the Royal Society in 1871, the Prix Poncelet of the Paris Academy of Sciences in 1870, and was lionized in his home town of Heilbronn (there is a statue of him there).

Joule eventually became a grandee of science, was awarded the Copley medal of the Royal Society in 1872, and has been honoured by having his name used to denote the modern unit of energy: 1 joule (J) is the work done by a force of 1 newton acting through a distance of 1 m. But his enduring image is that of someone who was merely a brilliant experimenter and accurate evaluator of the mechanical equivalent of heat. Having seen his work, we can now appreciate that he was more than just exceptionally deft and painstaking. He showed great physical intuition in his identification of obscuring effects and in his close reasoning. Quantities such as I^2R or the 'mechanical equivalent of heat' were not just waiting there to be dusted down and discovered. As Newton had said, with reference to his own '*experimentum crucis*' on colour, Nature has to be coaxed into revealing her secrets.

If the time was ripe, then why did it still take so long for Mayer and Joule's work to be recognized? There are many answers: both were outsiders to the scientific establishment; Joule further noted a North/South divide (in provincial Manchester, we 'have dinner at noon'[22]); Joule was not a charismatic speaker (he was shy, perhaps due to his being slightly hunchbacked); and Mayer made some egregious errors in his early work. We must also realize that we have hindsight—we now know that energy is important. In the 1840s there were other exciting trails to blaze. There was the discovery of Neptune by Adams and by Le Verrier, Hamilton's prediction of conical refraction, and Armstrong's discovery that a jet of high-pressure steam was electrified[23]—could electricity be generated this way? Heat physics was a bit *passé* and Fourier had already said the last word on it.

Crucially, there was also the 'category error'—heat and work were radically different sorts of things; and the heat-to-work conversion was particularly hard to demonstrate. Also, it seemed that in order to bring out the laws of *nature*, some very *unnatural* contrivances had to be employed: the electromagnetic machine was a very complicated device and bore about as much relation to dal Negro's prototype (a pendulum swinging near magnets) as Watt's steam engine did to Hero's 'kettle'. (The reason for this increase in complexity became apparent only after the arrival of yet another new concept—'entropy'; see Chapters 16 and 18.)

Mayer and Joule always referred everything back to heat-work conversions, but there is no doubt that they had inklings of the more abstract and comprehensive concept of *energy*. They insisted that the principle of the conservation of total 'force' (energy) would apply within the diverse domains of physiology, light, electricity, magnetism, 'living force' (kinetic energy), and 'force of attraction' (potential energy), as well as heat and mechanical work. (For example, Joule carried out 28 experiments on the heat-equivalent of light-generation in combustion.[24])

Mayer, however, took a positivist approach: when 'heat' is consumed, it simply ceases to exist and is replaced by a 'mechanical equivalent'. The numbers all come out right and there is no need to ask what heat actually is. Joule, on the other hand, took Dalton's atoms and the dynamical theory of heat as the bridge between the very different realms of work and heat. As work was motion, then heat must be a 'state of vibration' and not a 'substance'.[25] Joule formulated a proto-kinetic theory (he had already come across Herapath's work; see Chapter 11) and calculated that the speed of atoms in a gas, and also of 'water atoms',[26] was just over a mile per second. (This figure was also consistent with the very high speed of his incandescent meteors.)

Mayer and Joule never met or exchanged letters. Mayer had evaluated the mechanical equivalent of heat in 1842, Joule in 1843. Mayer quickly sent a letter to the *Comptes Rendues* in 1848 when he feared that his work had been overlooked. There was an acrimonious priority dispute, but this was conducted by third parties (Tyndall, Clausius, and Helmholtz for Mayer; Tait, Thomson, and Rankine for Joule). Mayer was admiring and respectful of Joule's work; Joule's private verdict was that he was 'content to leave Mayer in the enjoyment of having predicted the law of equivalence. But it would certainly be absurd to say he has established it.'[27] In a letter to Tyndall in 1862, he expanded on this: 'I think that in a case like that of the Equivalent of Heat the experimental worker rather than the mere logical reasoner (however valuable the latter) must be held as the establisher of a theory. I have determined the mechanical equivalent in nearly a dozen ways, and the figure I arrived at in 1849 has not yet been altered or corrected... Believe me, Dear Tyndall, Yours always truly, J.P. Joule.'[28]

The neglect of Waterston's work highlights the enormous contemporary difficulty in conceptualizing a gas as a 'swarm' of a near infinity of miniscule molecules moving at high speeds and—the greatest difficulty of all—in a random or chaotic fashion. The story is continued in the work of Clausius, Maxwell, and Boltzmann in Chapter 17, in the section on Kinetic Theory.

15

Faraday and Helmholtz

The time was indeed ripe for the discovery of energy, but this didn't mean that the path to it was obvious or that there was only one such path. Already (in the previous chapter), we have met those researchers whose main goal was to determine the exact equivalence value between heat and mechanical energy. Now (in the late 1840s), there were those who rather looked for qualitative unifying features in all the 'forces' or 'powers' of nature. One such was Michael Faraday. Finally, there was one researcher, Hermann von Helmholtz, who sought to set out the entire theoretical framework, deriving the actual formulae for energy in all its various forms, and who came up with the first definitive statement of the conservation of energy.

Michael Faraday

Michael Faraday (1791–1867) was born into a poor family in South London (he was once given a loaf of bread to last a week) and had only a very rudimentary education (reading, writing, and ciphering). At 13, he had a job delivering newspapers. His employer, a French émigré, Monsieur Ribeau, not only hired out newspapers but also sold and bound books, and thus it was that, at 14, Faraday became an apprentice bookbinder and, thereby, an avid reader. Two publications aroused his interest in science: an article on 'Electricity' in a copy of the *Encyclopaedia Britannica* that he was rebinding and Mrs Marcet's *Conversations in Chemistry*. Jane Marcet had written it after attending Sir Humphry Davy's lectures at the Royal Institution.

One day, a customer at Ribeau's offered Faraday some tickets for Davy's lectures. He eagerly attended the lectures, took careful notes, and bound these notes in a special volume. However, by October 1812, his apprenticeship had finished and he wrote: 'I must resign philosophy entirely to those who are more fortunate in the possession of time and means…I am at present in very low spirits.'[1]

But then an accident occurred that had great promise in it, for Faraday at any rate. Davy was temporarily blinded by an explosion with nitrogen trichloride (the same substance that had injured Dulong's eye and finger; see Chapter 10) and was recommended Faraday as amanuensis. This came to pass, but only lasted for a few days. In late December, a desperate Faraday wrote to Sir Humphry begging for employment and sending along the bound volume of carefully written out lecture notes. Davy was flattered but still couldn't help. Then, fortune smiled on Faraday for a second time. The 'fag and scrub' at the Royal Institution, a Mr Payne, lived up to his name and became involved in a brawl. He was summarily discharged: 'That evening…Faraday…was startled by a thundering knock at the door. On the street below he saw a carriage from which a footman alighted and left a note for him in which Sir Humphry requested him to call the next morning…'[2] The rest, as they say, is history.

Faraday came from a small religious sect, the Sandemanians, and to understand God's universe was the driving spirit behind all his ambitions in science. His friend, the physicist John Tyndall (see Chapter 14), observed, perplexedly, 'he [Faraday] drinks from a fount on Sunday which refreshes his soul for a week'.[3] For Faraday, as for Joule, God's 'powers' could not be created or destroyed without some compensatory balance. As Faraday was later to write: 'The highest law in physical sciences which our faculties permit us to perceive—[is] the Conservation of Force.'[4] But in addition, Faraday was convinced that there had to be an underlying *unity* in all the powers or forces: 'the various forms under which the forces of matter are made manifest have one common origin'.[5] To demonstrate this by experiment was the common thread behind the whole of Faraday's long career.

The scientific world was agog after Oersted's linkage between electricity and magnetism in 1820 (see Chapters 9 and 14). The young Faraday was one of many who immediately began to investigate this further, but Faraday brought his philosophy of the 'unity of force' to bear on it: if electricity caused a magnet to rotate, then surely magnetism would cause electric 'rotations'? His experiment to demonstrate this (in 1821) was one of the most ingenious experiments ever devised—on a par with Pascal's vacuum-in-a-vacuum (see Chapter 4)—and, significantly, brought out the symmetry between magnetism and electricity (Fig. 15.1) as well as the unity of force. Leaving aside the sorry tale of Humphry Davy's accusations of plagiarism, Faraday's predictions were borne out—the free wire (with cork attached) rotated around the fixed magnet when the battery was connected.

André-Marie Ampère (1775–1836) (sometimes dubbed the 'French Faraday') had recently shown that electric currents exhibited magnetism. Faraday therefore predicted that magnetic 'currents' would produce electricity. This led to the most famous of all Faraday's discoveries, his law of electromagnetic induction. But he showed this in *two* totally different ways: first, by a varying magnetic 'field' (as we should say now) and, secondly, by a moving magnet. Feynman and Einstein were struck by this and Feynman wrote: 'We know of no other place in physics where such a simple and accurate general principle requires for its real understanding an analysis in terms of *two different phenomena*.'[6]

Less memorable but also important were Faraday's researches (in around 1832) showing that all the various kinds of electricity (whether from electrostatic generators, voltaic cells, thermocouples, dynamos, or electric fishes and eels) were identical. In the Fifteenth Series of Experimental Researches, on the 'character and direction of the electric

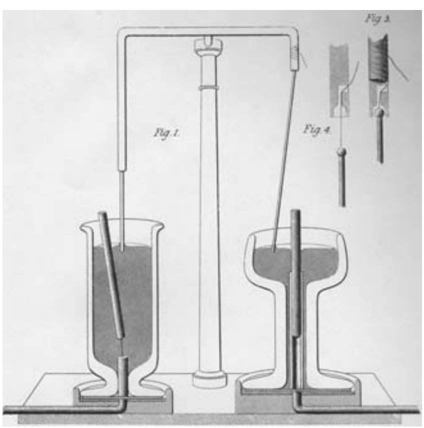

Fig. 15.1 Faraday's electromagnetic rotations, from *Experimental Researches in Electricity*, volume 2, 1844 (on the left, a magnet is free to rotate through mercury round the fixed conductor; on the right, the magnet is fixed and the conductor rotates freely).

force in the *Gymnotus* [electric eel]', he explicitly stated that the convertibility of force included *all* manifestations:

> Seebeck taught us how to commute heat into electricity; and Peltier . . . how to convert the electricity into heat. Oersted showed how we were to convert electric into magnetic forces, and I had the delight of adding the other member of the full relation . . . converting magnetic into electric forces. [Now I have further shown that the electric eel can] convert nervous into electric force.[7]

This was followed by Faraday's researches into electrolysis (in 1833), which showed the links between electricity and chemical affinity: the amounts of different substances deposited or dissolved by the same quantity of electricity were proportional to their chemical equivalent weights. This law of electrochemistry was Faraday's only *quantitative* law. Surprisingly, it didn't lead Faraday on to accept atoms: 'if we adopt the atomic theory . . . then the atoms of bodies which are equivalents to each other . . . have equal quantities of electricity . . . But I must confess I am jealous of the term *atom*; for though it is very easy to talk of atoms, it is very difficult to form a clear idea of their nature.'[8] While his law of electrochemistry didn't lead him on to atomism, it did throw doubt on the contact theory of electricity.

The contact theory had been promulgated by Alessandro Volta (1745–1827), inventor of the voltaic pile. He had argued that electromotive force was generated by the mere contact of dissimilar metals; no chemical action was required. This smacked of perpetual motion—a metal could be brought up to another metal any number of times, generating electric power but without consumption of anything. Faraday, in 1839, carried out a series of experiments demonstrating unequivocally that chemical action was *always* present, although one might have to look hard to find it. But Faraday's objections to the contact theory were metaphysical as well as experimental:

> By the great argument that no power can ever be evolved without the consumption of an equal amount of the same or some other power, there is *no creation of power*; but contact would be such a creation.[9]

This was as close as Faraday ever came to a statement of the conservation of energy.[10]

Still searching for evidence of the 'unity of force', Faraday, in August 1845, resumed experiments that he had started some 25 years earlier. Polarized light was passed through an electrolytic cell and the plane of polarization was checked for rotation after various alterations to the external conditions. Faraday tried different electrolytes (distilled water, sulphate of soda solution, sulphuric acid, and copper sulphate) and different currents (constant, intermittent, beginning, ceasing, and rapidly recurring secondary currents)—but the light was not rotated. He later wrote to Sir John Herschel, 'It was only the very strongest conviction that Light, Mag[netism], and Electricity must be connected that could have led me to resume the subject and persevere through much labour before I found the key.'[11]

In September, Faraday tried again with electromagnets instead of galvanic currents and transparent materials instead of electrolytes. The polarized light was passed through flint glass, rock crystal, and calcareous spar and the magnetic 'field' strength was varied—but still no rotation was observed. Then Faraday tried using some glass that he had made for the Royal Society back in 1830, a 'silico borate of lead', a glass with an extremely high refractive index. At last, '*there was an effect produced on the ray*, and thus magnetic force and light were proved to have relation to each other'.[12] (He further went on to show that he had discovered a new kind of magnetism, which he named 'diamagnetism' and which, with characteristic thoroughness, he tested on everything from glass to foolscap paper, from litharge to raw meat.[13])

We have already stated Faraday's misgivings about atoms. Influenced by his former patron, Davy, who had been influenced in his turn by Coleridge and Kant, Faraday veered rather towards point-atoms and 'force fields' similar to those advocated by Boscovich (see Chapter 7, Part II). (These views were not popular—Gay-Lussac and Louis Jacques Thénard even threatened police action if one of Davy's papers was published.[14]) The force fields had the advantages of being continuous and extendable to infinity. More germane was the fact that they ruled out the need for imponderable subtle fluids. This could only encourage the new, emerging concept of energy—for how could energy be conserved if heat and electric and magnetic fluids were all *separately* conserved?

By the late 1840s and 1850s, Faraday had further evolved his worldview in a unique and revolutionary way—powers were disseminated by three-dimensional 'lines of force'. These lines of force filled all space and accounted for the unity of force and the harmony of the cosmos. It was all part of the 'Grand Design' where 'nothing is superfluous'.[15]

There was only one effect left to consider and that was gravity—surely it was linked to electricity. Faraday's own powers were waning and this was to be his last piece of research. (Faraday had a breakdown in 1838–40, probably due to overwork. After 1860, he could do no more research.)

Gravity troubled him greatly: a body at a height of several thousand feet weighed less than one on the surface of the Earth, but where (Faraday asked) was the compensatory force accounting for the loss of weight? 'Surely this force must be capable of an experimental relation to Electricity, Magnetism and the other forces, so as to bind it up with them in reciprocal action and equivalent effect',[16] he wrote in 1849.

Experiment would be the guide, and these investigations would perhaps be the most important of Faraday's whole career, putting his cherished principle of the unity of force to the test: 'It was almost with a feeling of awe that I went to work.'[17]

Cylinders of copper, bismuth, iron, gutta-percha, and so on, fell from the ceiling to a 'soft cushion' on the floor in the lecture room at the Royal Institution. The cylinders were surrounded by a helix of copper wire, 350 feet long, connected to a sensitive galvanometer[18]—but no effect was found. In the next experiments, the cylinders were vibrated rapidly up and down—still no effect, but 'They do not shake my strong feeling of the existence of a relation between gravity and electricity.'[19] Finally, the whole experiment was scaled up. A 280-lb weight of lead was raised and lowered through 165 feet within the Shot Tower on the River Thames near Waterloo Bridge. Again, the results were negative, but Faraday insisted 'I cannot accept them as conclusive.'[20] He wanted to repeat them with more sensitive instruments, but this was not to be. This was Faraday's last paper submitted for publication (and rejected), in the spring of 1860.

Faraday recognized that gravity was strange in at least two respects: it had no neutral state (whenever there was mass, then there was gravity) and it was much weaker than the other forces: 'a grain [about 0.002 oz] of water is known to have electric effects equivalent to a very powerful flash of lightning'.[21] He continued, 'many considerations urge my mind toward the idea of a cause of gravity which is not resident in the particles of matter merely, but constantly in them *and all space*' (emphasis added). With hindsight, it seems as if Faraday was on the trail of a *field* theory of gravity...

This would be consistent with Faraday's overall outlook. It was the intuitive concept of force rather than the more abstract concept, energy, that fired his scientific imagination. He was the discoverer of lines of force in electricity and magnetism and was a tireless researcher into the 'unity of force'. Even though his discoveries had led directly to the electric motor and the dynamo, the final linkage—that between the various forces and mechanical *energy*—doesn't appear to have commanded his attention. Energetic quantities such as 'mechanical equivalent', 'work', and 'gravitational potential' are directionless and may be tracked by lines, like the contour lines on an Ordnance Survey map: but these lines give only the height, they don't show which way a pebble will actually roll.

All along, Faraday seems to have been questing for a unified field theory, a quest that continues to this day, and in which gravity still refuses to join the party. Thus Faraday, like Newton, missed 'energy'. We now turn to our last scientific personage, von Helmholtz, who, when he wrote 'force', really did mean 'energy' (most of the time).

Hermann von Helmholtz

Hermann von Helmholtz (1821–94) (the 'von' was added by Kaiser Wilhelm I in 1882) was what one might call a physicists' physicist: he made outstanding contributions in both experimental and theoretical physics, was 'exceptionally calm and reserved',[22] and was a patriarch of German science and a physics professor from 1871 until his death in 1894. It comes as something of a surprise, therefore, to find out that his early researches were in physiology and, worse, his first and most influential paper in physics started out with a long introduction on philosophy.

Helmholtz was born in Potsdam, Germany, the eldest of four children. His father was a teacher at the Potsdam Gymnasium and a highly cultured man. (He was a personal friend of the philosopher, Immanuel Fichte, son of *the* philosopher, Gottlieb Fichte.) The young Hermann was accomplished in the arts as well as the sciences. In one of his first letters home when a student in Berlin, he writes: 'Any spare time I have during the day is devoted to music…I play sonatas of Mozart and Beethoven…In the evenings I have been reading Goethe and Byron…and sometimes for a change the integral calculus.'[23] His father advised him: 'don't let your taste for the solid inspiration of German and classical music be vitiated by the sparkle and dash of the new Italian extravagances—these are only a distraction, the other is an education'.[24]

Helmholtz's first love was physics, but his father's salary didn't extend to university fees and, besides, 'physics was not considered a profession at which one could make a living'.[25] He readily accepted medicine as an alternative, as the fees were paid in return for working for some years as a Prussian army doctor after graduation.

Within medicine, Helmholtz veered towards physiology but saw every problem through physicist's eyes. (He taught himself mathematics and physics through private study in his 'spare time'.[26]) The vitalist philosophy of Stahl (also inventor of the phlogiston theory; see Chapters 5, 6, and 8) and the '*Naturphilosophie*' of Hegel held sway in Germany at this time. The vitalists attributed life in organisms to the presence of a 'life force' in addition to food, air, and water. Helmholtz felt that this was 'contrary to nature',[27] but was unable to state his misgivings in the form of a definite question.

Finally, in his last year as a medical student, Helmholtz 'realized that Stahl's theory treated every living body as a *perpetuum mobile*'.[28] But he had known, ever since his teenage years, that perpetual motion was supposed to be impossible, and this was reinforced by his recent reading of the works of Daniel Bernoulli, Euler, d'Alembert, and Lagrange (as usual, in 'spare moments'). He now had an aim—to rid physiology of vitalism. He was joined in this quest by three other young physiologists in Berlin, especially his friend and fellow student, Paul du Bois-Reymond.

Thus Helmholtz's first researches, between 1845 and 1848 (having recently qualified as a doctor), were all intended to rebut the vitalists' claims. Specifically, he carried out experiments to try to show that the 'mechanical force and the heat produced in an organism could result entirely from its own metabolism'.[29]

This was easier to state than to demonstrate—we all know, for example, that some people can eat like a horse and others like a sparrow. Some of the problems that Helmholtz had to contend with were as follows: he had no value for the mechanical equivalent of heat; heat could be generated in the blood and muscles as well as in the lungs and stomach; heat of excreta had to be accounted for; also what were the relevant heats of combustion of the food products and had the reactions gone to completion?

Helmholtz estimated that 2.6% of heat was lost to excrement, 2.6% to heating expired air, 14.7% by evaporation from the lungs, and 80.1% by evaporation of sweat and radiation and conduction from the skin.[30] He realized that the previous studies (of Dulong and Despretz) were in error in using the heats of combustion of hydrogen and carbon rather than of the complicated molecules of fat or sugar. He also invoked Hess' Law to show that the order of the decomposition reactions didn't matter.

Helmholtz had still to show that food accounted for the mechanical work done by the animal. He continued with experiments on the thigh of a frog (he 'waited impatiently for the spring and the frogs'[31]) and found that a single muscular contraction caused a temperature rise of 0.001–0.005°C. He improved or invented instrumentation as he went along: for example, the myograph, whereby a contracting muscle leaves its trace on the blackened surface of a revolving cylinder or on a moving glass plate; the 'moist chamber', for keeping the muscle in a good condition; and electrical apparatus to apply shocks of known duration and intensity. (His most famous invention was the opthalmoscope in 1851, and others included the Helmholtz resonator and Helmholtz coils.) All this research did show that vital forces were redundant, but the experiments could not be considered as absolutely conclusive.

Then, in 1847, while still a Prussian army surgeon with the Royal Hussars, Helmholtz had a change of tack. The motivation was still to show that perpetual motion was impossible, but he now widened the scope to cover all physical processes, with physiology just a subset, and also moved from an experimental to a theoretical attack. He had a grand ambition: he wanted to base his new principle of the impossibility of perpetual motion on the securest possible foundation—on philosophical bedrock, as it were—and to apply it to the whole of physics.

With remarkable assurance (still only 26, this was his first work in physics, a subject in which he was entirely self-taught) he composed a memoir, some 60 pages long, entitled '*Über der Erhaltung der Kraft*' ('On the conservation of force').[32] It started off: 'I have [formed] a physical hypothesis...and then developed the consequences of this hypothesis in the several branches of physics and, finally, have compared these consequences with the empirical laws.'[33]

For the justification of his ideas, Helmholtz looked to Kant's transcendental idealism. There he found that there were two main kinds of law in science: empirical laws (a summary of observations—the laws of refraction of light and Boyle's Law were examples) and theoretical laws (laws of the hidden causes behind the observations). But which kind of law was the prohibition against perpetual motion? We have seen that perpetual motion was vetoed because of the *experimental* findings—people kept trying, but the machines always failed (see Chapter 2). However, Helmholtz, like Mayer, Carnot, and others, thought that the principle was rather of the second kind—it just *had* to be true.

From Kant again, science had to be comprehensible and its comprehensibility lay in the law of causality. Some things might happen spontaneously or freely—that is, without cause—but they were outside the remit of science. Helmholtz took the impossibility of perpetual motion as an example of a causal law and, in fact, equivalent to the old nostrum of 'cause equals effect'.

Did the impossibility of perpetual motion mean that something was being conserved?

Kant's philosophy was helpful yet again. It asserted that the physical world was comprehensible because of two 'intuitions'—space and time—and two abstractions, matter and force. Matter was inert and could only be experienced through its effects. But—and this was the crucial link—force was the *cause* of all the effects. Thus, Helmholtz was able to associate 'cause' with 'force' and 'cause equals effect' with 'conservation of force'. In short, the impossibility of perpetual motion *implied the conservation of 'force'*.

Finally, what if several different causes could all explain the observations equally well? Helmholtz argued that only the *simplest* (i.e. those satisfying the principle of sufficient reason) could be considered as the 'objective truth'. Now Kant's 'matter' had only two properties—mass (i.e. 'intensity') and relative location. For Helmholtz, it was evident, therefore, that the simplest type of forces were central—dependent only on the masses and relative position of the bodies, and not on their orientations, velocities, or accelerations. Moreover, the forces acted only along the line joining the two bodies; that is, they were all either attractions or repulsions.

After a lengthy philosophical introduction in which these antecedents were explained, Helmholtz applied his new principle, 'the conservation of force', across all branches of physics.

First, he turned to the most well-established discipline, that of mechanics. Now Sadi Carnot's syllogistic argument limiting the maximum efficiency of heat-engines (Chapter 12) was based, ultimately, on the impossibility of perpetual motion. Helmholtz was very impressed by Carnot's argument (as Clausius and Thomson soon would be—see Chapter 16—and, indeed, as we are today). So he imported Carnot's argument, along with Carnot's new constructs of states and cycles, from 'heat physics' into mechanics. He considered a system of bodies acted upon only by central, mutual (i.e. internal) forces, F, and he defined work as $\int F\,dr$. A given state was now defined by the positions and velocities of all the bodies at a given time. The work released in changing the system from an initial state, A, to a final state, B, had to equal the work required to take the system from B back to A, or 'we should have built a *perpetuum mobile*'.[34]

In this way, Helmholtz was able to prove all the standard results in mechanics: 'conservation of *vis viva*'; that the maximum quantity of work in going from A to B was definite and fixed (shades of Carnot); the principle of virtual velocities; that a system of bodies can never be set in motion by the action of its internal forces; that the final speed in free fall depends only on the vertical distance travelled; that this speed is just sufficient to return the body to its starting height (assuming no friction); that the work done in a simple machine is inversely proportional to the speed of its moving parts; and the laws of elastic impact (when combined with the constancy of the motion of the centre of gravity).

Considering next the case of wave motion, Helmholtz noted that the intensity of a wave had to decrease in accordance with the inverse square law. Fresnel, Helmholtz observed, had already shown that the laws of reflection, refraction, and polarization of light at the boundary between two media could be deduced from the conservation of *vis viva* and continuity requirements. Furthermore, interference between two wave trains resulted only in a redistribution of intensity—so the total 'force' was conserved. However, experiment had still to confirm that the total heat radiated by a body A was equal to that absorbed by the body B.

Helmholtz commented that the impact of inelastic bodies and friction were two processes in which 'an absolute loss of force has until now been taken for granted'.[35] However, for inelastic impact, the tensional forces due to deformations were increased and heat and sound were also generated. Friction always led to an equivalent amount of thermal and electrical changes.

In the conversions from mechanical work to heat, Helmholtz mentioned Joule as being the only investigator (he had not yet come across Mayer), but he was not particularly complimentary about Joule's methods: 'His [Joule's] methods of measurement ... meet the difficulties of the investigation so imperfectly that the results can lay little claim to accuracy ... a quantity of heat might readily have escaped ... loss of mechanical force in other parts of the machine was not taken into account.'[36] Helmholtz again mentioned Joule's work when it came to the generation of heat from magneto electricity, endorsing all of Joule's careful deductions against the materiality of heat (see Chapter 14), but neglecting to credit Joule with them.

Regarding the conversion of heat to work, Helmholtz commented that 'nobody has yet bothered', but then went on to cite the very experiments where Joule had done just that (Joule's experiments on the adiabatic expansion of gases). At least this time Helmholtz admitted that the experiments had been 'rather carefully made'.[37] All in all, one gets the impression of a young man (Helmholtz was 26) impatient to have his ideas experimentally corroborated, rather than of any serious criticism of Joule. (In the revised edition of the memoir in 1881, Helmholtz was more generous towards Joule: 'His [Joule's] later investigations, carried out with complete professional knowledge and indefatigable energy, merit the highest praise.'[38])

From all the contemporary evidence, Helmholtz was convinced of the wrongness of the caloric theory. He attributed heat instead to the *vis viva* of the constituent particles ('free heat') and to any tensional forces between these particles ('latent heat'). He also acknowledged Ampère's theories whereby the *vis viva* could manifest itself as rotational as well as translational molecular motion. With the dynamic theory of heat, Helmholtz asserted that the 'principle of the conservation of force holds good wherever the conservation of caloric was previously assumed'.[39]

Electricity and electromagnetism took up a third of Helmholtz's memoir. This was an area in which the Germans were especially active, and Helmholtz was fully up to date with the researches of Gauss, Weber, Franz and Carl Neumann (father and son), Lenz, Ohm, Kirchoff, and others.

In static electricity the forces were central and balanced the 'living forces'—the principle of the conservation of force was guaranteed. Helmholtz defined the equipotential surfaces of an isolated conductor and outlined operations for establishing the unit of electrical potential. He also gave the (now) standard expression for the energy of a capacitor, $\frac{1}{2}Q^2/C$, and compared it with experiments measuring the heat generated by the discharge of a Leyden jar.

In galvanic (i.e. current) electricity, Helmholtz straightaway dismissed Volta's contact theory: 'the principle which we are presenting here directly contradicts the earlier idea of a contact force'.[40] Instead, the contact force should be replaced by (central) forces of attraction and repulsion 'which the metallic particles at the place of contact exert upon the electricities at that point'.[41]

Helmholtz went on to use his conservation principle to deduce the 'electromotive force' of a battery by balancing this 'force' against the heat generated chemically in the cell and the resistive heating in the wire. He brought to bear all the relevant, contemporary, quantitative laws: Ohm's Law, Lenz's heating law, Joule's more general heating law, Kirchhoff's circuitry laws, and Faraday's laws of electrolysis.[42]

He then derived another heat balance equation for the case of thermoelectric currents (those arising out of junctions held at different temperatures, as in the Peltier effect—see the end of Chapter 9), and bemoaned the fact that there were no quantitative experiments with which to compare his predictions.

Electrodynamic and induction phenomena were concerned with the motion of magnets near currents in wires. Lenz's Law and Franz Neumann's extension of it showed that 'the force of the induction current . . . acts always in opposition to the force which moves the magnet'.[43] This was encouraging—a perpetual motion was thereby prevented.

Carl Neumann had defined an 'electromotive force of induction' and had an expression for the work done by this force in moving the magnet. Helmholtz used Neumann's result to show that 'if a magnet moves under the influence of a current, the *vis viva* which it gains must be supplied by the tensional forces consumed in the current'.[44] (The heating in the wires, I^2R, had also to be remembered.) He was further able to show that Neumann's undetermined, empirical constant of proportionality was the reciprocal of the mechanical equivalent of heat.

However, Helmholtz's wording in the memoir suggested that all this followed from his conservation principle alone. Neumann and also Clausius protested and Helmholtz eventually capitulated. For one thing, Helmholtz had not taken into account self-induction; for another, he had implicitly assumed that the 'energy' of the circuit-magnet system didn't depend on the position of the magnet (in fact, it didn't, but this wasn't self-evident[45]).

These examples show that the case of electromagnetism was tricky. In fact, it was worse than tricky, it was the first instance in which Helmholtz's programme failed—the forces were no longer all of the simple, central type that Helmholtz had assumed. Weber had traced back electrodynamic and induction phenomena to 'the forces of attraction and repulsion of the electric fluids themselves',[46] but the intensity of these forces was found to depend upon 'the [magnet's] velocity of approach or recession and upon the increase in this velocity'.[47] Helmholtz commented dryly that 'Up to the present time no hypothesis has been found by which these phenomena can be reduced to constant central forces.'[48] However, the major part of his project, the conservation of force, *was* still upheld.

Finally, Helmholtz returned, very briefly, to the initial impetus for all his work—the absence of vital forces in organic processes. He said that for plants 'there is a vast quantity of chemical tensional forces stored up . . . the equivalent of which we obtain as heat when they are burned'. Also, 'the only *vis viva* which we know to be absorbed during the growth of plants is that of the chemical rays of sunlight; we are totally at a loss, however, for means of comparing the force equivalents which are thereby lost and gained'.[49]

For animals, he said that 'Animals take in oxygen and the complicated oxidizable compounds [food] which are generated in plants, and give them back partly burned as carbonic acid and water, partly reduced to simpler compounds. Thus they use up a certain quantity of chemical tensional forces and . . . generate heat and mechanical force.'[50] From the calorimetric experiments of Dulong and Despretz, the validity of the conservation of force can 'at least approximately, be answered in the affirmative'.[51]

Helmholtz wrote in conclusion:

> I believe, by what I have presented in the preceding pages, that I have proved that the law under consideration [the conservation of force] does not contradict any known fact within the natural sciences; on the contrary, in a great many cases it is corroborated in a striking manner. . . . I have attempted to guard against purely hypothetical considerations, [and] to lay before physicists as fully as possible the theoretical, practical, and heuristic importance of this law, the complete corroboration of which must be regarded as one of the principle problems of physics in the immediate future.[52]

Reception of the Memoir

This was a bravura performance: Helmholtz had founded the 'conservation of force' on the securest of foundations, had shown an extraordinary knowledge of contemporary theory and experiment, and the work of the ancients, and had identified, or had a good stab at identifying, the appropriate formulae—the 'blocks of energy'—in fields as disparate as heat, electricity, magnetism, physiology, and mechanics. In outlining a future programme for physics, he had shown enormous self-confidence, not to say chutzpah. But what did his contemporaries make of it?

The older physiologists (still followers of Stahl) were not to be persuaded by an exposition that was so mathematical, theoretical, and in a foreign discipline.

Helmholtz had high hopes of impressing the physicists, at least, but to his 'astonishment' the physicists 'were inclined to deny the correctness of the law [of the conservation of force] . . . because of the heated fight in which they were engaged against Hegel's philosophy of nature, [and] to treat my essay as a fantastic piece of speculation'.[53] What irony, that Helmholtz, in trying to banish Hegel's *Naturphilosophie* from science, should have come over as too Hegelian. The scientist and publisher Poggendorff (who had rejected one of Mayer's papers) declined to publish Helmholtz's work as it was too theoretical, too speculative (philosophical), and, most damning of all, had no new experimental findings. As regards theoreticians, only one, Carl Jacobi (Chapter 13), was enthusiastic, while Franz Neumann was mildly supportive, Weber indifferent, and Clausius not at all receptive.[54]

It was really only the British school—in particular, William Thomson (Chapter 16)—that gave whole-hearted support to Helmholtz's memoir. Thomson came upon Helmholtz's 'admirable treatise'[55] in 1852 and wrote that had he 'been acquainted with it in time',[56] he would have used its results in many of his papers. Helmholtz and Thomson first met in 1855 and became close friends for life. They evolved to have somewhat similar outlooks, both considering the law of the conservation of energy as the heart of physics. They also grew to occupy similar positions as establishment figures of science in their respective countries. (In the priority dispute over Mayer and Joule, Helmholtz, Thomson, and all the other physicists except for Tyndall split along national lines, but this didn't sour international physics relations for long.)

James Clerk Maxwell was also appreciative of Helmholtz's memoir, writing (in 1877):

> To appreciate the full scientific value of Helmholtz's little essay on this subject [the conservation of energy], we should have to ask those to whom we owe the greatest discoveries in thermodynamics and other branches of modern physics, how many times they have read it over and over, and how often during their researches they felt the weighty statements of Helmholtz acting on their minds like an irresistible driving-power.[57]

(Helmholtz was the first to promote Maxwell's electromagnetic theory on the continent. He subsequently developed his own theory, which included Maxwell's as a limiting case.)

It is fascinating to learn of contemporary life and the interrelations between these energy physicists. Helmholtz met Faraday on his first trip to England in 1853. He wrote to his wife: 'I succeeded in finding the first physicist in England and Europe . . . Those were splendid moments. He is as simple, charming, and unaffected as a child; I have never seen a man with such winning ways. He was, moreover, extremely kind, and showed me all there was to see. That, indeed, was little enough, for a few wires and some old bits of wood and iron seem to serve him for the greatest discoveries.'[58]

Helmholtz met Thomson in Kreuznach, a German resort, in 1855: 'I expected to find the man, who is one of the first mathematical physicists of Europe, somewhat older than myself, and was not a little astonished when a very juvenile and exceedingly fair youth, who looked quite girlish, came forward. He is at Kreuznach for his wife's health . . . she is a charming and intellectual lady, but in very bad health. He far exceeds all the great men of science with whom I have made personal acquaintance, in intelligence and lucidity and mobility of thought, so that I felt quite wooden beside him sometimes.'[59] He met Thomson again in Glasgow in 1864 and Thomson's brother, a professor in engineering, was also there: 'It is really comic to see how both brothers talk at one another, and neither listens, and each holds forth about quite different matters.'[60]

On this trip, Helmholtz also met Joule at a dinner party: 'Mr Joule, a brewer and the chief discoverer of the conservation of energy, and [another guest] were both very pleasant lively individuals, so we spent a most interesting evening.'[61] Finally, he also met Maxwell: 'I went with an old Berlin friend to Kensington, to see Professor Maxwell, the physicist at King's College, a keen mathematician, who showed me some fine apparatus for the Theory of Colours which I used to work at; he had invited a colour-blind colleague, on whom we experimented.'[62]

Of London, Helmholtz wrote to his wife, 'And now you shall hear about this great Babylon. Berlin, both in size and civilization, is a village compared to London';[63] and in humorous vein (after talking at the BA meeting at Hull), 'Here in England the ladies seem to be very well up in science . . . they are attentive and don't go to sleep, even under provocation.'[64]

Overview

First, a comment on words. In the birth of a new concept, as with the birth of a child, there is usually an interim period during which an appropriate name must be found. In the memoir Helmholtz used only the term 'force', sometimes in the old Newtonian way but usually in the new way, meaning 'energy'. It is invariably clear from the context which meaning is intended; thus 'forces of attraction and repulsion' (force means force) and 'the principle of the conservation of force' and 'force equivalent' (force means energy). The term 'energy' was introduced by Thomson in 1852 and Helmholtz very much approved.

History can sometimes miss the point: Joule is often regarded as having 'merely' carried out extremely precise measurements and Helmholtz as determining lots of formulae, as if working back from the correct dimensions of energy. So, what new did Helmholtz's memoir bring in? Maxwell, giant of physics in the nineteenth century, says it all in his appraisal of Helmholtz:

> [Helmholtz] is not a philosopher in the exclusive sense, as Kant, Hegel . . . are philosophers, but one who prosecutes physics and physiology, and acquires therein not only skill in discovering any desideratum, but wisdom to know what are the desiderata.[65] (1862)

and

> the scientific importance of [Helmholtz's] principle of the conservation of energy does not depend merely on its accuracy as a statement of fact, . . . it [further] gives us a scheme by which we may arrange the facts of any physical science as instances of the transformation of energy from one form to another. (1877)[66]

Helmholtz's memoir marked the beginning of a new era, the 'epoch of energy'.[67] However, there was a big difference in the philosophical underpinning of the German and British schools (this will become evident in Chapters 16 and 17). Thomson, Joule, Faraday, and to a lesser extent Maxwell invoked the permanence of 'God's Creations' while Mayer, Clausius, and Helmholtz argued from the rationality of science.

There can be no doubt that it was Helmholtz's expertise across disciplines (the three 'phs' of physics, physiology, and philosophy, and also mathematics, the 'science of aesthetics', and the 'moral sciences') that gave him such an extraordinarily broad outlook. Each discipline helped the others along, as we shall see.

From his work in physiological optics and physiological acoustics, Helmholtz learned that there was no such thing as raw sense data. Fichte's view was that we only experience a succession of conscious states—but this doesn't permit us to know if there is really anything 'out there'. Helmholtz cleverly adapted Fichte: our free will permits us to contrive experiments that lead to correlations between these experiments and our conscious states—thus there *is* something out there. (Helmholtz waited until after the death of his father—a Kantian Fichtean idealist—before expounding these views.)

According to Helmholtz, the correlations are captured in the laws of physics, and these laws grew to have more significance than causes for him. (As Helmholtz's friend, du Bois-Reymond, commented, nothing is to be gained by introducing causes such as the hand 'that shoves the inert matter silently before itself'.[68] In other ways, also, Helmholtz progressed on from Kant. For example, his work on the origins of geometry convinced him that space doesn't have to be Euclidean. But in the final analysis, and by Helmholtz's own reckoning, he always remained a Kantian at heart and never departed from Kant's dichotomy of 'existence' and 'activity', from which Helmholtz derived 'matter' and 'energy'. This is still the prevailing conceptual divide in physics today (cf. the physicist explaining his subject at the start of this book).

One can take philosophy too far. Helmholtz had wanted to prove not only that total energy was conserved, but also that elemental force was *necessarily* central. Clausius, in 1853, disagreed. Helmholtz derived the force between two bodies from the sum, over all pairs, of the force between a mass-point in one body and a mass-point in the other body. Furthermore, according to the Principle of Sufficient Reason (simplicity), the force between mass-points could only depend on the distance between them (after all, by Kant, the only variable property of matter was its position). Clausius disagreed again—perhaps nature isn't so simple and the force between two mass-points *does* depend on other factors such as direction. (We shall discuss the thermodynamics of Clausius in the next chapter.)

In the case of electromagnetic forces, nature decided in favour of Clausius rather than Helmholtz. The relative orientation of bodies (e.g. magnets and circuit elements) *is* important—in other words, the whole is more than the sum of its mass-points. Worse still, Weber had shown that in electrodynamic phenomena the force depended on speed and acceleration (e.g. on *how fast* a magnet was moved).

Helmholtz conceded that 'Up to the present time no hypothesis has been found by which these phenomena can be reduced to constant central forces.'[69] He took hope from the fact that Weber's force-law led to instabilities and sometimes to perpetual motion and must therefore be flawed. But Clausius, with his amazing clarity and logic, found that if Newton's Third Law ('Action equals Reaction') was sacrificed, then perpetual motion was avoided and the conservation of energy upheld even with velocity-dependent forces. Helmholtz, with reluctance, had to agree. He added an appendix to the 1881 edition of his memoir in which he stated that the forces were central only in those special cases where the Principle of the equality of Action and Reaction also applied.[70]

There was yet one more problem with Helmholtz's philosophy. Helmholtz, still influenced by Kant, could never sanction forces or energy in empty space. His pupil, Heinrich Hertz (discoverer of Maxwell's electromagnetic waves), commented that Helmholtz always considered that matter was the seat of force and that without matter there was no force. But Faraday's lines of force were *curved*, even where there were no mass-points along the way. Even stranger, sometimes the lines of force were closed loops, neither starting nor ending on a massy source. Finally, Maxwell's equations (1873) were to show that the electromagnetic field had energy, even in 'empty' space (see Chapter 16).

Helmholtz struggled to accommodate the new physics into his worldview. He developed an electrodynamics that was a bridge between the Continental action-at-a-distance theories and Maxwell's theory (in fact, Maxwell's theory was a limiting case of Helmholtz's). He went on to apply thermodynamics to chemical processes and he defined the 'free energy' of a chemical system as the energy that is available for conversion into work. He then used the Second Law to derive an equation (the Helmholtz–Gibbs equation) showing that it is the free energy rather than the heat that is maximized. This explained the seemingly anomalous occurrence of spontaneous endothermic reactions (those chemical reactions in which net heat is absorbed).

However, because of all the philosophical problems, Helmholtz, while never abandoning his cherished law of the conservation of energy, moved away from seeing it as the pre-eminent guiding principle of physics. Instead, he pinned his hopes on the principle of least action. He spent the last few years of his life working on this, but with no very great success.

After the work of Helmholtz, 'vitalism' disappeared and the Germans, reluctant at first, came to adopt the energy principle with such enthusiasm that some even advocated energy as the primary quantity, in place of mass or force. These 'energeticists' (see Chapter 17) even went so far as to deny the existence of atoms—remarkable in view of the fact that most were chemists.

16

The Laws of Thermodynamics: Thomson and Clausius

The final players in our history of energy are Thomson and Clausius. Their seminal contributions were written in and around 1850, but neither made any reference to Helmholtz's memoir of 1847. Instead, they looked back some 25 years to Sadi Carnot's *Réflexions* (see Chapter 12). They initially had access only to Clapeyron's revision and extension of this work, but they both peered through Clapeyron to the genius of Carnot.

William Thomson

William Thomson (1824–1907) was the most gifted in a gifted family—his father and elder brother were professors of mathematics and engineering in their time. Born in Belfast, William's mother died when he was six years old and the family moved to Glasgow, where his father became professor of mathematics at the university. Taught initially at home, William's mathematical prowess showed up early; aged only 16, he had mastered Fourier's *Analytical Theory of Heat* and shown Professor Philip Kelland of the University of Edinburgh to be in error regarding the Fourier series. (Fourier's work was to be an influence for the whole of Thomson's life—see Chapters 11 and 17.) Thomson studied mathematics at Cambridge (1841–5) and was reputedly so confident of success that he asked who came second in the mathematics tripos exams only to find that *he* had come second. He revised better for the next lot of exams and came first, as Smith's Prizeman, in 1845. We have already heard—in Chapter 13—how he tracked down Green's work at this time.

After Cambridge, Thomson went to Paris for a year, to work in Regnault's laboratory and learn some practical skills. Victor Regnault (1810–78) was a master experimentalist and was researching the thermal properties of gases, especially steam, and the gas laws and precision thermometry in general. Thomson then returned to the University of Glasgow as professor, a position he held from the young age of 22 until his retirement.

Fourier's work had introduced Thomson to heat studies and now at Regnault's laboratory he learned of the difficulties inherent in practical thermometry. Then Thomson came across Clapeyron's 1834 paper 'On the Motive Power of Heat',[1] a reworking of Sadi Carnot's *Reflections on the Motive Power of Fire* (see Chapter 12). It inspired Thomson to find Carnot's original work, which he managed after some effort (see Chapter 12). Thomson was greatly impressed by Carnot's *Reflections* ('the most important book in science'[2]) and saw what neither Clapeyron himself nor anyone else had seen—that Carnot's theory opened the door to the possibility of another scale of temperature apart from, say, the height of a column of mercury or the volume or pressure of a gas. According to Carnot's extraordinary proof, the maximum efficiency of a heat-engine didn't depend on the engine design or working substance, but depended only on the actual operating temperatures (see Chapter 12). Therefore, Carnot's Law could yield a temperature scale that was *absolute*—independent of any device or material. This was the holy grail of temperature scales, once and for all removing questions regarding the compatibility and uniformity of different types of thermometer.

After a first paper in 1848 and then another in 1854, Thomson finally arrived at *the* temperature scale that is still in use today and called the Kelvin scale (Thomson was elevated to the peerage in 1892, becoming Lord Kelvin, Baron of Largs).

The Kelvin Scale of Temperature

The scale can now be explained as follows (where Q refers to the heat transferred within a Carnot cycle). Imagine lots of little ideal heat-engines (Carnot cycles) taking in heat at a high temperature, T_{high}, and giving heat out at a lower temperature, T_{low}. Now it is evident that $Q(T_{high})$ must be greater than $Q(T_{low})$ as $Q(T_{low})$ is the heat *left over* after work has been done. (We remember from Chapter 12 that this asymmetry is endemic in Carnot's engines.) Also, T_{high} is defined as being higher than T_{low} from Carnot's directive that the heat-engine goes in the forward direction—in other words, produces motive power—by transferring heat from higher to lower temperatures.

We have the further requirement that all the miniature engines operating between any two reservoirs must be of the same 'size'—they must be scaled to each other before an overall scale can be set up. This scaling can be achieved, hypothetically, at any rate, by making sure that all the engines give out the same heat at some reference reservoir.

These two provisos—that T_{high} is defined as greater than T_{low} and that all the engines are scaled—are all that is required to define the absolute scale of temperature. What results is remarkably simple: T is proportional to Q_T. In other words, the temperature of any given heat reservoir, T, is determined by the heat, Q_T, that would be taken in or put out at that reservoir by a pre-scaled ideal heat-engine. It is not even necessary to ask what has or will happen to that heat—whether it shall be conducted, radiated, converted to/from work, or whatever.

The efficiency of an ideal heat-engine operating between T_1 and T_2 is defined as W/Q_1, where W is the work done. From the First Law of Thermodynamics, $W = Q_1 - Q_2$, and therefore the efficiency is $(Q_1 - Q_2)/Q_1$. Using the absolute temperature scale, we can now rewrite this as follows:

$$\text{Efficiency}=(T_1-T_2)/T_1 \tag{16.1}$$

As the efficiency can never be greater than unity, so the temperature can never be less than zero. There is therefore an absolute zero to the temperature.

We can now calculate the absolute efficiency of any ideal engine provided that we know the temperatures. For example, Carnot's test case of an alcohol-vapour engine running between 78.7°C and 77.7°C has an ideal efficiency of (78.7 – 77.7)/(273.15 + 78.7) or 0.28%—not a very encouraging result. Real engines, like cars etc., have a higher efficiency due to a greater starting temperature and drop in temperature.

Conflicting Theories

While developing the absolute scale of temperature, Thomson still (in 1849) hadn't come round to accepting the conversion between heat and work. He thought that Carnot's conclusions would come crashing down if the axiom of the conservation of heat was abandoned. However, he couldn't reconcile this axiom with Joule's experiments, which showed that heat and work were interconvertible and that heat was actually consumed or generated in the process (see Chapter 14).

One problem, in particular, that troubled Thomson was the phenomenon of heat conduction. In this process, heat was transferred from a high temperature, say at one end of an iron bar, to a lower temperature at the other end of the bar (he may have been thinking of Fourier's conduction equation, yet again)—but where was the work that, according to Carnot, should have been developed by this 'fall' of heat? Thomson wrote:

> When thermal agency is thus spent in conducting heat through a solid what becomes of the mechanical effect [work] which it might produce? Nothing can be lost in the operations of nature—no energy can be destroyed.[3]

This is the first recorded use (in 1849) of Thomson employing the word *energy*.

Another similar problem was the case of a swinging plumb-line (or pendulum), which loses all its motion when submerged in water—what has become of the 'mechanical effect'? Thomson, after Joule, carried out his own experiments with a paddle-wheel and found that, like Rumford, he could bring water to the boil merely by friction. Thomson had thus amply confirmed Joule's findings—but this experiment still didn't suggest to Thomson the solution to his conflicting theories.

Yet another problem concerned the hypothetical case of an engine based on the freezing of water: as the water froze and expanded it could do work (e.g. cause a pipe to burst). But this all occurred at a constant temperature (32°F) and so contradicted Carnot's requirement for a fall of temperature. Thomson's brother James came up with an ingenious resolution: as work was done and the pressure on the ice increased, then the temperature of melting might be lowered. William Thomson immediately set about testing this suggestion with some very careful experiments; he managed to confirm James' prediction and give precise quantitative support to Carnot's theory.

It seemed as if both Carnot's and Joule's opposing theories were becoming more and more confirmed and entrenched. Thomson couldn't resolve this conflict; nor could he explain the conundrum of workless heat conduction, and the fact that Joule appeared to have shown conversions of work to heat but not of heat into work. (Thomson was a bit unfair to Joule over this, as Joule *had* shown the generation of work from heat by the adiabatic expansion of a gas against an external pressure; see Chapter 14.)

Salvation came from Rudolf Clausius, with his amazing clear-sightedness and cool logic. He saw that Joule's interconversions had to be correct and that Carnot's conservation of heat had to be jettisoned. But he found a way of achieving this without putting Carnot's Law in jeopardy.

Rudolf Clausius

Rudolf Clausius (1822–88) was born in Köslin, Pomerania (now Poland). His father was a Lutheran pastor and principal of a small private school at which Rudolf and his many older brothers had their early education. He then went on to a Gymnasium in Stettin, to the University of Berlin (1840–44) and to present his doctoral dissertation in 1847 at the University of Halle.

There is a three-year gap (1847–50) during which little is known about his life, and then comes the publication of his famous paper, 'On the moving force of heat and the laws of heat that may be deduced therefrom',[4] in 1850.

The importance of this work was soon recognized in Germany and in England (thanks to Tyndall's translations) and Clausius' academic career was launched—he was successively a professor in Berlin, Zurich, Würzburg, and finally Bonn, from 1869 until his death in 1888. He became one of the leading physicists (the others were Helmholtz and Kirchhoff) in the newly unified Germany, a nation that was to become pre-eminent in science and industry during the second half of the nineteenth century.

There is scant background information about Clausius. He married Adelheid Rimpam in 1862 and volunteered to lead an ambulance corps in the Franco-Prussian war of 1870–71. He was wounded in the leg and awarded the Iron Cross for bravery, but suffered severe pain and disability for the rest of his life. (In this same war, Regnault's laboratory was destroyed and his son killed.) Clausius' wife died in childbirth in 1875 and Clausius was left to look after their six children, a task that he apparently took on with dedication and kindness. In 1886 he married again and had one more child. We make further brief mention of Clausius' work regarding the kinetic theory in Chapter 17.

The First Law of Thermodynamics

We often hear it said that the Second Law was discovered before the First, but this wasn't quite the case. Clausius was convinced by Joule's heat⇋work conversions (the reversible arrow means both heat-to-work and work-to-heat conversions) and saw that these were ultimately explained by the dynamical theory of heat—that heat is a motion of the microscopic constituents. He was nevertheless conscientious in developing the First Law (his 'First Principle of Heat') without recourse to any extraneous microscopic assumptions. How he did this was as follows.

Clausius re-analysed Carnot's cycle with just one radically different premise: whenever work was done, heat was actually consumed (not just gone latent) and whenever work was consumed, then heat was generated (not just released from a latent form). He tackled Carnot's scenario of an expanding gas (see Chapter 12). (Also, he later reworked Clapeyron's example of a vapour in contact with its liquid.) Carnot's analysis had been hampered by the fact that he couldn't track the heat all the way round the cycle. Clausius got round this by considering an infinitesimal cycle, thereby making the curves in the P versus V graph (see Fig. 12.2) approximate to straight lines. In effect, the cycle was now a straight-sided figure, a tiny parallelogram, and more amenable to calculation and experimental support. The heat taken up along the horizontal isothermal sections could be found using Mariotte's Law and the vertical constant-volume sections checked against the data on C_V. Clausius found—as he had expected—that Carnot's heat-conservation axiom was not upheld, but that the work done, W (the area of the parallelogram), plus the heat transferred, Q, *together* summed to a constant quantity, ΔU:

$$\Delta U = Q + W \tag{16.2}$$

(Also new and important, Q and W were both in Joule's 'mechanical units'.) This was to become the First Law of Thermodynamics—but what was U?

It was clear to Clausius that Q and W were not separately conserved and, worse, were dependent on the route taken around a Carnot cycle. However, Clausius saw that the change in U, ΔU, *was* constant between any two states and, moreover, could be expressed as a function of the parameters that defined those states (the temperature and the volume). Thus U was a 'function of state' and was determined solely by the properties intrinsic to the given state. U was, in fact, the intrinsic or internal *energy* of the state. It was given this designation later by Thomson, but Clausius was the first one to appreciate that more was implied than Mayer's and Joule's conversions between heat and work—a new abstraction, *energy*, was needed.

To summarize, Clausius' First Principle of Heat, or the First Law of Thermodynamics, went beyond Mayer's and Joule's heat⇋work conversions and posited a new entity, the total internal energy of a system. For an isolated system, or for a state returned to, the total internal energy was a constant, and, unlike Q or W, could be expressed as a function of the intrinsic state variables, V and T, or P and T, and so on.

Thus far, Clausius' arguments were quite general and macroscopic; but he separately hypothesized that U was the total of all the internal microscopic motions (the 'internal free heat') and the total of all the intermolecular potential energies (the total 'internal work'). The free heat and the internal work depended on detailed microscopic assumptions particular to the given system; they could not, in general, be translated into a functional relationship of macroscopic state variables, P, V, or T. Ironically, Clausius' analysis had been for the rather special case of heat added to an expanding ideal gas. Here there are no intermolecular forces, so the internal work is zero and Q does then equal the change in internal heat, which, in turn, is equal to ΔU. Thus, in this special case, the internal heat *is* a function of state (it's equal to $C_V\,\mathrm{d}t$, where $\mathrm{d}t$ is the temperature interval).

The Second Law of Thermodynamics

After this establishment of the First Law came Clausius' great resolution of Thomson's problem—the conflict between Carnot and Joule. For work to be generated, Joule required the *consumption* and Carnot the *transfer* of heat—but Clausius saw that there was still the possibility that *both* of these processes occurred. Moreover, it could be the case that the amount of work done depended on the *proportional split* between these processes (heat consumed and heat transferred), and that this in turn depended only on the initial and final temperatures. In other words, Carnot's Law could still be upheld.

Clausius wrote in his famous work of 1850, 'On the moving force of heat and the laws regarding the nature of heat itself which are deducible therefrom':[5]

> it may very well be the case that at the same time [as work is produced] a certain quantity of heat is consumed and another quantity transferred from a hotter to a colder body, and [that] both quantities of heat stand in a definite relation to the work that is done.

In order to determine quantitatively what this 'definite relation' was, Clausius realized that, as usual, the extremum case had to be examined, in other words, he had to look at the maximum work done by ideal engines. Therefore Clausius again analysed infinitesimal Carnot cycles, now incorporating the First Law, and compared his findings with the calculations of Clapeyron and of Thomson, also using the extensive data of Regnault. His results agreed with these earlier analyses tolerably well (there was some departure from Clapeyron in the cases where the vapours were of high density). All the findings corroborated Carnot's and also Joule's suspicion that the ideal efficiency was not only temperature-dependent but actually varied as $1/T$.

But how did the familiar form of the Second Law come out of this? It is a testimony to Carnot that the Second Law arrived, not from any calculations or comparisons with data, but from another run-through of Carnot's original syllogistic argument (see Chapter 12 again). This time, Clausius employed a subtly different starting premise from Carnot.

Clausius imagined one heat-engine (say, 'superior') that could do more work than an ideal heat-engine, or, 'what comes to the same thing'[6], that it could do the same amount of work for a smaller amount of *transferred* heat. Running the ideal engine in reverse and following it with the superior engine going forward, the work consumed would be totally compensated for, but there would be some residual heat transferred to the higher temperature. The net result would be a new combined engine that did no work but transmitted heat from the cold to the hot reservoir.

Clausius didn't like this result and so ruled out the possibility of a 'superior' engine. Why didn't he like this result? It certainly didn't contravene the First Law, as the net work done was zero and all the heat was conserved; all that had happened was that the heat was distributed differently, transferred from the lower to the higher temperature. Clausius' objection was simply that it didn't tie in with common experience—heat *doesn't* flow from cold to hot of its own accord, that is, without the input of work.

In the same way that Carnot had ruled out a 'superior' engine because it would permit a perpetual source of motive power, Clausius vetoed it because it would allow heat to flow from cold to hot unaided. In other words, both men advanced physics by appealing to age-old wisdom and common experience: perpetually acting machines and heat flowing 'uphill' don't happen—the world just isn't like that. This statement of Clausius', that heat cannot flow from a low temperature to a higher temperature unless aided by work, was the first appearance of the Second Law of Thermodynamics (Clausius' 'Second Principle of Heat'[7]).

Thomson immediately saw the force of Clausius' arguments, agreed with his conclusions, and saw that his own conflict had been resolved. Also, Thomson had at last (during 1850) come to accept Joule's results and the dynamical theory of heat. He wrote a paper in 1851, swiftly following Clausius' paper of 1850, in which he acknowledged Clausius' priority but put Clausius' statement into a more precise form.

Statements of the Second Law

It is impossible for a self-acting machine, unaided by external agency, to convey heat from one body to another at a higher temperature.[8]

Thomson also came up with his own statement:

It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects.

Here, he unknowingly addressed the query in Carnot's posthumous notes (see Chapter 12) as to whether 'motive power could be created... by mere destruction of the heat of bodies'.[9] According to Thomson this can't be done—you can't simply cool the ocean or the atmosphere and convert the extracted heat into work.

Either of these statements—that due to Clausius or that due to Thomson—constitutes a statement of the Second Law of Thermodynamics. Thomson stated (and it can easily be shown) that they are completely equivalent to each other (the violation of one implies the violation of the other).

Finally, in 1854, Thomson put the Second Law into a more quantitative form.[10] From his absolute temperature scale (defined in this same paper—a paper mostly on thermoelectricity), he already knew that Q was proportional to T (referring, as always, to the operations of ideal heat-engines). Then, in a moment of profound perception, Thomson realized that this was tantamount to a mathematical expression of the Second Law. Adopting the convention that heat taken out of a reservoir was positive and heat put in was negative, then—however complex the cycle, in other words however many different temperature reservoirs were involved—the sum of Q/T for a complete reversible cycle must always come to zero:

$$\sum(Q/T) = 0 \qquad (16.3)$$

(complete reversible cycle)

Cosmic Generalizations

So far, the Second Law applied only to heat-engines. It was a prohibition against heat flowing spontaneously from cold to hot without the aid of work, and, equivalently, a prohibition against cooling a body below the temperature of its surroundings and converting all that heat into work. It was Thomson who, seemingly out of the blue, extended the canvas of the Second Law to—all of nature…

How this happened is more or less as follows (there is some conjecture in this). Implicit in the Second Law is the assumption that heat *does* flow spontaneously from hot to cold. Thomson saw that such allowed heat-flows not only occurred readily in nature but gave an overall direction to natural processes: hot things cool down; heat flows from the hot end of a bar to the other end; heat flows from the Sun to the Earth; and so on.

He then saw a commonality between these examples and other processes where there was a natural direction in time but which had nothing overtly to do with heat: the pendulum slowing down in water, planets slowing down over aeons, frictional losses in a mechanical device, and so on. The unifying feature for all these multifarious phenomena, on both terrestrial and planetary scales, was a loss of useful 'mechanical effect', whether actual (the pendulum stops swinging) or potential (the work that didn't happen during the conduction of heat).

Another unifying feature was that all these processes gave a cosmic direction to time—they were irreversible. Now what were the links between heat flowing, loss of mechanical effect, and irreversibility?

By 1851, Thomson had understood it all. In the first place, Carnot's Law only applied to ideal *reversible* processes and heat conduction was clearly irreversible. Secondly, Carnot only imposed a maximum and not a minimum to the amount of work generated: the minimum might be zero, as in the case of heat conduction. Finally, Thomson saw that energy conservation, which he had at last accepted, was not contradicted by irreversible processes.

In switching from the caloric theory to 'energy', Thomson had also to switch from the material to the dynamic theory of heat. With this new acceptance—of the dynamical *microscopic* theory of heat—Thomson realized that bulk motion lost was made up for by an increase in the microscopic motions. (We remember, from Chapter 3, Leibniz's example of large money converted into small change.) Some mechanical effect (a bulk phenomenon) was lost to mankind, but it was never lost *in toto* (when all the microscopic motions were taken into account).

Some common irreversible processes were heat conduction, friction, and the absorption of light. All were *dissipative*. This was a word from the Victorian lexicon (Thomson's father had apparently told him cautionary tales about idle young men dissipating their talents[11]). It was dissipation that seemed to Thomson to lie behind all the disparate phenomena (in heat conduction, friction, and the absorption of light, bulk energy had been dissipated into increased microscopic motions). Then, in 1852, all these ideas came together in three bold declarations, stated abruptly at the end of a brief paper entitled 'On a Universal Tendency in Nature to the Dissipation of Mechanical Energy':[12]

> There is at present in the material world a universal tendency to the dissipation of mechanical energy…
>
> Any *restoration* of mechanical energy, without more than an equivalent of dissipation, is impossible…
>
> Within a finite period of time past, the earth must have been, and within a finite period of time to come the earth must again be, unfit for the habitation of man…

As usual, it is hard to be sufficiently impressed by the boldness and sweep of these assertions. We cannot easily un-know what now seems so obvious and well-attested: that things age and decay, that cosmic time has a direction. We shall discuss at the end of the chapter the impact of Thomson's words and the validity of such generalizations. We shall find that dissipation wasn't the end of the story and wasn't the clinching characteristic that dictated the direction of time. The next leap forward in understanding came with Clausius.

Entropy

Clausius had realized the necessity of a new abstraction (energy), resolved Thomson's conflict between Joule and Carnot, discovered the Second Law, checked up that Carnot's ideal engine efficiency did indeed vary as $1/T$, and made great strides into the kinetic theory (in 1857; see Chapter 17). All this would have been enough to guarantee his place as a founder of thermodynamics—but he pushed back the frontiers even further (in the 1850s and 1860s), striving to get to the very heart of the Second Law.

Energy conservation ensured that the 'energy books' balanced—that for every bit of heat that was consumed an equivalent amount of work was generated and vice versa. Was there, Clausius wondered, *another* conservation law and *another* abstract quantity that determined the *proportions* of heat converted into work and heat transferred to another temperature?

As usual, quantification would only come by looking at the extremal case, the case of the reversible ideal heat-engine. In this ideal heat-engine, Clausius saw that there were basically just two kinds of 'transformations', the conversion of heat into work and the transfer of heat. Clausius wondered whether there was an equivalence between these two transformations; after all, they both concerned heat and they always occurred together in Carnot's cycle. Heat did not flow spontaneously from cold to hot, but when this process did occur—say, in a reversed ideal engine—it was because it was always *compensated* for by a simultaneous conversion of work into heat. Presumably there was some 'equivalence value' for each type of transformation that guaranteed this compensation.

What could the quantitative form of the 'equivalence value' be? Clausius had an amazing ability to see past distractions through to the interior logic, to see the general in the particular.

First, Clausius saw that Carnot's quintessential heat-engine involved, most generally, *three* temperatures and not just two. In the most general case, heat was transferred between, say, temperatures T_x and T_y, and work was generated from heat at *yet another* temperature—say, T_z.

Secondly, he saw that the transfer of heat was really the same as two other processes: a conversion of heat into work at the higher temperature and a conversion from work back into heat at the lower temperature. In other words, all transformations were just of *one* type, heat⇋work, and it was not necessary to ask whether the work itself was of the 'shifting weights' or the 'shifting heat' variety.

Lastly, even this one type of transformation, heat⇋work, was itself only an example of heat added or subtracted from a heat reservoir, never mind about the work. Because heat and work were quantitatively connected (by the First Law), all the heat magnitudes were taken care of—they would all come out in the wash, so to speak.

The money-laundering analogy can be followed a little way. We have seen in Chapter 14 that the 'currency exchange rate' (Joule's mechanical equivalent of heat) is fixed even though the proportions may change (the *bureau de change* may see a long queue and to avoid running out of notes change only 20% of the money proffered). Now we are further finding that the money or heat has no trace of its origins, whether stolen or clean, whether obtained from heat-transferred or from work. But now we must ditch this analogy. In money transactions, the money that can be withdrawn or added to a bank is not related systematically to that bank. However, in physics, the value of the heat *is* dependent on the temperature of the heat bank: the higher the temperature, the less valuable is a given quantity of heat.

Carnot had defined the worth of heat in terms of how much work it could do. Now Clausius was seeking to generalize this notion: in essence, his 'equivalence value' was a more abstract measure of the worth of heat.

The 'equivalence value' that Clausius was looking for had to have the mathematical form $Q/f(t)$, as it depended only on the heat, Q, and on some increasing function, f, of the temperature, t. (An increasing function is, as its name suggests, a function that always has larger values for increasing values of t.) In fact, as $f(t)$ is always increasing, it might just as well be *defined* to be equal to the absolute temperature, T. Thus Clausius finally arrived at the following formulation: within a Carnot cycle, the equivalence value for every addition or subtraction of heat, Q, at temperature T, is Q/T.

(Clausius was undoubtedly influenced by Thomson's mathematical formulation of the Second Law; see Equation 16.3.)

There was just one thing left to do, and that was to specify a convention for the sign of the heat. Clausius decided that wherever a process was natural, spontaneous, then it should be given a positive sign. So adding heat (as if converted from work) should be positive and with-drawing heat (as if to convert it into work) must be negative. (There was still the question of whether heat was added to the system or to the reservoir—the surroundings—and it must be admitted that Clausius wasn't always consistent in this.)

Clausius' original aim had been to show that something is being conserved. In going around a complete Carnot cycle, the system is brought back to its initial state and nothing has changed. Clausius interpreted this as the requirement that the sum of the equivalence values for a complete cycle must be zero. Generalizing to infinitesimal transfers of heat, $\mathrm{d}Q$, Clausius finally arrived at

$$\int \mathrm{d}Q/T = 0 \tag{16.4}$$

This was plainly the conserved quantity that Clausius had been searching for. He labelled it S and, in his paper of 1865, gave it the name entropy:

$$S = \int \mathrm{d}Q/T \tag{16.5}$$

He especially constructed a word that sounded similar to 'energy' and was likewise rooted in classicism:

> I hold it better to borrow terms for important magnitudes from the ancient languages so that they may be adopted unchanged in all modern languages, I propose to call the magnitude, *S*, the *entropy* of the body, from the Greek word τροπη, transformation.[13]

To recapitulate, Clausius was claiming that, in a complete Carnot cycle, the total change in entropy was zero. It was entropy and energy, rather than caloric, that were conserved in the running of the ideal heat-engine. This was an outstanding advance: however, as it stood, the conservation of entropy was a somewhat barren result as, after all, it didn't apply to any practical engines or, indeed, to anything real at all.

It was Clausius alone among his contemporaries who saw how to make the result universal and applicable to everyday phenomena. The Carnot cycle, while having important universal aspects—such as not being limited to a particular engine-type or substance—was still limited to reversible processes. Clausius realized that for irreversible processes the entropy changes did not scatter equally on either side of zero but were all clumped together on one side.

In these real irreversible processes (everything that actually happens), there was always *less* than the ideal maximum of work generated. Less work meant more heat put out at the lower temperature. This, Clausius saw, translated into an integral over dQ/T *exceeding* zero in all real processes (we are talking now of heat added to the surroundings). Thus, combining the real and the ideal cases, this meant that in *all* changes the total entropy either stayed the same or increased. Clausius summarized this as follows:

$$S = \int dQ/T \geq 0 \tag{16.6}$$

(The symbol ≥ is read 'greater than or equal to'.) This came to be known as another way of stating the Second Law of Thermodynamics.

But what was entropy? Consideration of this question caused Clausius such difficulties that he delayed publication of his work for over a decade (the relevant papers came out in 1854, and then 1862 and 1865). The definition of entropy had been arrived at by a paring away of elements until there was nothing left but Q and T. By this 'sleight of maths' the quantitative relations between heat, temperature, and entropy had emerged. Clausius nevertheless realized that in order to gain insight into the physical nature of entropy it was essential to put the 'work' and molecular considerations back in. (Up to this point, Clausius had been admirably careful to keep his macroscopic theories separate from his microscopic speculations. We must remember that the existence of atoms wasn't confirmed until the twentieth century.)

When heat was added to a system, Clausius saw that any or all of the following might happen: an increase in the internal 'free heat'; an overall change in physical dimensions (usually an expansion); some other changes in structure (e.g. melting, evaporation, dissolving, chemical reaction, etc.). The first corresponded to an increase in the temperature or average speed of 'thermal' molecular motions, and all the other effects amounted to changes in molecular arrangement of one sort or another. These changes in molecular arrangement (the 'disgregation'[14]) were all examples of 'work', done to or by the system. This link to molecular arrangements implied that entropy was an *extensive* quantity—in other words, it had extension just like length, breadth, mass, or volume (see footnote at end of Chapter 13).

Two other properties that Clausius was able to establish were that, first, like U, the entropy was a function of state and, secondly, neither U nor S could be determined absolutely; that is, they could be determined only up to some arbitrary zero-point. (In Equations 16.5 and 16.6, the zero-point has been assumed to be zero.)

Clausius finished off this paper (in 1865), like Thomson's paper of 1852, with some sweeping cosmic assertions:

The energy of the universe is a constant; E = constant (16.7a)

The entropy of the universe tends to a maximum; $\Delta S \geq 0$ (16.7b)

Overview

Two new concepts have emerged, energy and entropy, and two universal laws, the first governing the conservation of energy in all its forms, and the second concerning the distribution of thermal energy. (That Equation 16.7b determines the *distribution* of energy will be explained in Chapter 18, in the section on entropy.)

It is impossible to say who contributed more to these discoveries out of Thomson and Clausius. Thomson was more influential in introducing 'energy' to the whole of physics, whereas Clausius' legacy was particularly in the area of thermodynamics. It was Clausius who realized that the subtle heat-fluid had to be replaced by a subtle concept, energy.

All along, from the late 1840s onwards, the torch of discovery was passed back and forth between Thomson and Clausius, the work of each one crucially shaping the work of the other. This happened through their publications—it is not known whether they ever met. They each had one other physicist closer to hand (William Rankine and Helmholtz, respectively), but Clausius barely gave Helmholtz's work a mention—it was on energy in all its forms rather than on thermodynamics—and Rankine, although he made important contributions, was criticized for not keeping separate his new theories in thermodynamics and his speculative molecular theories.

There was nevertheless a difference in style between the British (Thomson, Rankine, and Joule) and the Germans (Clausius, Helmholtz, and Mayer). The former appealed to religious convictions (there were many references to 'the Creator' and quotes from the Bible) while the latter couched their theories in neutral terms. However, for all, the theories themselves were ultimately judged by purely physical aesthetic criteria.

It is something of a puzzle as to why Thomson was the last to finally come around to accepting energy, the more so as Joule was his close friend and associate. The following must all have had something to do with it.

The enormous success of the caloric theory and the fudge-factor of latent heat continued to operate right into the middle of the nineteenth century. The works of Clapeyron, Thomson, and Regnault, within the caloric theory, produced a wealth of data that was not immediately overturned by Clausius' almost identical analysis of Carnot cycles. (Clausius' analysis, of course, had the signal difference that he invoked energy and the First Law.)

The very fact of the Second Law also made the First Law harder to spot (the 'energy books' didn't balance exactly in practice, and also heat → work conversions were not so evident as work → heat conversions).

Thomson had been won over by Fourier at an early age. Thomson's biographers make the interesting point that Fourier's heat conduction theory and Thomson's theories of electricity bore many parallels.[15] During a reversed Carnot cycle, heat could be 'raised to a higher state'. In the same way, work could be done to move electricity to a new state (to a higher electrical potential). Electricity was conserved in this process and so, Thomson presumed, heat was also conserved in like process.

Although Thomson was the last to accept energy, he then proceeded apace, with all the fervour of the recent convert. It was he who introduced the terms 'energy' and 'thermodynamics' into physics (in 1851) and then expanded 'energy' to cover all applications, not merely the interaction of heat and work. In 1852 he adopted Rankine's terms 'potential' and 'actual' energy and, subsequently (1862), substituted the term 'kinetic' for 'actual' energy. With fellow Scot, Peter Guthrie Tait (1831–1901), he wrote what was to become the standard textbook in physics, *Treatise in Natural Philosophy*,[16] and so launched what his biographers refer to as the 'epoch of energy'.[17] Despite this, Thomson stuck with a force-based Newtonian outlook, as opposed to endorsing Hamilton's variational mechanics where energy takes centre stage. The variational approach was given more prominence on the Continent and led to a British/Continental divide that continues to this day.

Thomson was also emerging as an establishment figure. He amassed a personal fortune as a result of his work developing the first transatlantic telegraphic cable and was a rising star in the Victorian scientific firmament. It had been completely out of character for Clausius to put forward his cosmic generalizations (Equations 16.7a and 16.7b), and one wonders whether he felt compelled to compete or was simply inspired by Thomson's forthright style and public persona. Their relationship seems to have been one of mutual respect but no particular warmth, and they were in opposing camps as regards the priority dispute between Mayer and Joule. Clausius also had to contend with his concept of entropy being (wilfully?) misunderstood by Tait, a topic we return to in the next chapter.

The expression $S \geq 0$ is the first time that an inequality comes into use in physics (some later examples are Heisenberg's Uncertainty Principle and the limiting speed of light; see Chapter 17). This, along with Thomson's assertions on dissipation, introduces an overall direction to time. It is hard to overstate the impact of this on physics. Up until Thomson and Clausius, the legacy of Newton and of rational mechanics was that the universe was cyclical and might continue on forever in the same form. (The question of whether it had always existed in the past was, in Europe, answered in the negative for religious reasons.) Even when Laplace, in his *Celestial Mechanics*, found that frictional and tidal forces would make the planets and their moons gradually acquire circular orbits, he nevertheless assumed that once this had been achieved then the orbits would continue unchanged, forever.

The cooling of bodies was, as we have seen (Chapters 6 and 11), well known to Black and to Fourier, but their emphasis was on achieving equilibrium rather than on direction. Water found its own level, but who noticed the connection between this and the fact that rivers always flow down hill? Fourier's heat-conduction equation did have a direction for time built into it, but it seems that Fourier never noticed this.

The waste and inefficiency of engines was known about and lamented by engineers the world over. Also, it was known that there was a loss of motion in inelastic collisions and the flight of projectiles. There was also the conversion of wood into ash and gases in burning; the disappearance of salt and sugar in dissolving; the heating of electrical wires in a circuit; the ageing of people and things; the cooling of bodies; and so on. But, before Thomson and Clausius, these various phenomena were not connected with each other or to any other body of theory.

One puzzle with thermodynamics is: what have engines and their efficiency, so anthropic, got to do with physics? The extension of the empirical base from heat-engines to *everything* happened as follows. First, Thomson and Clausius realized that any real mechanism is like a heat-engine of sorts—some irreversible losses always occur. Then, it was understood that the losses are not always *heat* losses (e.g. in reality, Dennis' blocks don't just cool down, they wear down as well—and they also get dispersed as in Feynman's allegory). This loss or dissipation gives an overall direction in time (the blocks wear down, not up). In other words, it is dissipation rather than cooling that is the more general process.

But nature is subtle, and if one looks even more closely one can find processes that do have a natural direction in time but don't involve cooling or even obvious dissipation (in the sense of things getting lost or wasted). One such is that puzzling phenomenon, looked at in Chapter 10, the 'free expansion' of a gas. Here, 'free' means that the ideal gas does no work as it expands. The gas follows a given isothermal curve (Fig. 10.3), but always in the direction towards larger volumes and lower pressures. No work is done, no heat transfers take place, and nothing is lost or even getting more disorganized. What is driving this progression?

Clausius was able to show that as the expansion proceeds, the average molecular separation increases and so the 'disgregation' and hence the entropy increase. Clausius thus demonstrated that entropy was an extensive quantity, albeit of an abstract sort, and that entropy is the one property that always increases.

But why does this happen, why must entropy increase? This can't be answered until a fully microscopic and, above all, a statistical approach is adopted. All is explained in Chapters 17 and 18.

Microscopic explanations were also crucial to the acceptance of energy and the First Law of Thermodynamics. For example, they explained where the energy had gone to when the paddle-wheel, projectile, or pendulum had slowed down and come to rest: the bulk motion had been converted into an exactly equivalent amount of microscopic motions, in accordance with the dynamic theory of heat and of friction. (Incidentally, if Thomson could have looked even more closely at his 'paradoxical' pendulum, he would have found that it never comes completely to rest, but has a tiny residual *thermal* motion—see Chapter 18.)

The laws of thermodynamics stand in relation to Regnault's experimental work in the same way as Kepler's theories of planetary motion stand in relation to Tycho Brahe's measurements. However, in saying that the Second Law is empirically based, we mean that it stemmed from the wealth of 'common experience' as much as from detailed measurements.

Newton's Second Law of Motion, by contrast, is not primarily empirical, but arises from Newton's introduction and definition of new concepts—mass and force. The law of the conservation of energy appears as an intermediary case, based on both experience and ideology—perpetually acting machines don't exist and can't exist.

The Second Law, in the original statements of Clauisus and Thomson, was strangely wordy—not like usual laws of physics. (The Second Law wasn't put into a purely mathematical form until Carathéodory in the twentieth century; see Chapter 18.) What did contemporary physicists and society in general make of them? Thomson's ideas of dissipation, waste, and a future Earth not fit for the habitation of man could have engendered a deep despondency, or been as shocking as Kepler's non-circular orbits or Galileo's discovery of a pitted lunar surface. We discuss this further in Chapter 17, but in the main it seems that there was very little public response.[18] Perhaps the ideas were simply too big.

At this stage (the first decade or so after mid-century), energy had really 'come in' to physics. There were even those, the German 'energeticists', who later advocated that energy was *the* important concept, superseding force, mass, and even atoms. This therefore ends our historical chronology of the discovery of energy. In the next chapter, we look at interactions between the nineteenth-century physicists, how their ideas were modified and incorporated into modern physics, and how 'energy' came into the public domain.

17

A Forward Look

Even when an idea has 'come in', it doesn't mean that it's all plain sailing thereafter. For example, it took many years for the spread of the concepts of work and the potential function. Initially (the beginning of the nineteenth century), work was just for the 'workers' (i.e. the engineers, such as Coriolis, Navier, and Poncelet),[1] while the potential function was for the mathematicians (Laplace, Green, Poisson, Gauss, and others).

In circuit electricity, an interesting new feature was that mere topology (the connections in a circuit diagram) was not enough—the actual spatial distribution of wires, magnets, and so on made a difference. This was evidence that the system contained potential energy, the energy of *configuration*. However, the overlap between the energy measures VIt, $\frac{1}{2}CV^2$, and so on (where V is the 'voltage') and Green's potential function, confusingly also called V, wasn't at all obvious. Helmholtz and even Maxwell made mistakes, such as getting the sign of V wrong, counting it twice, or forgetting about it altogether.[2]

A partial answer is that VIt and $\frac{1}{2}CV^2$ (where V is the voltage, C is the capacitance, I is the current, and t is time) are tallies, measures of how much work is done in moving electric charge 'uphill' through the circuit; but what connection do these tallies have with Green's three-dimensional function, $V(r)$, or with Maxwell's newly emerging concept of the electrical energy-in-the-field?

In Lagrange's mechanics, we saw how $(T - V)$ contained directional information (e.g. the bead twists and turns along the given wire) even while the T and V themselves are scalars (see the end of Chapter 7). Likewise, in circuit electricity, the model contains spatial information implicitly; for example, the C might refer to a parallel-plate capacitor—an actual physical thing with a structure in three dimensions. Eventually, it came to be realized that integrating a volume element of the square of theelectric field over the relevant region in space yielded the *same* electrical energy as the tallies VIt or $\frac{1}{2}CV^2$. An analogy is an irrigation system where we want to know the total amount of water. We can employ a moisture-detector and track down every tiny volume-element of water in three dimensions; or we can add the number of standard components of known volume: 'length of pipe', 'holding tank', and 'cistern'.

There was even an additional, new complication. The circuit had potential energy, yes, but it was no longer of the simple kind encountered in Lagrange's mechanics: it depended on spatial configuration but also on velocities and on time. For example, the *rate* of cutting of flux-lines was found to be important.

Both *V*s were, in any case, soon to be superseded by Maxwell's concept of the energy-in-the-field (see below).

Kinetic Theory

The new concept of energy implied acceptance of the dynamic theory of heat and this in turn led to a new round in the tangled tale of the kinetic theory of gases. We keep on covering this in different chapters, but this is because it kept on being started afresh. So far, we have had Daniel Bernoulli in 1738, Cavendish in around 1787, Herapath in 1816, and Waterston in 1845. Now Clausius, in 1857, and with no knowledge of these precursors, derived his own kinetic theory (we have skipped over Krönig in 1856).

The chief innovation that Clausius introduced was the important concept of the 'mean free path'. This is the average distance that a molecule travels in a straight line in between collisions. In air at the usual pressure and temperature, this is of the order of 10^{-7} m. With the aid of this concept, Clausius could then answer the criticism of the Dutch physicist. Buys-Ballot as to why gas molecules moving at 0.5 km s^{-1} nevertheless cause only a slow diffusion of the smell of dinner across the dining room—they travel fast, but *not very far* between collisions.

James Clerk Maxwell (1831–79) developed a kinetic theory for gases (1867) in which he took a great leap forward: he considered that the molecules had a distribution of speeds rather than just one fixed speed, and—even more momentous—that this distribution was probabilistic. Gillispie[3] suggests that Maxwell was influenced by John Herschel's review of Quetelet's *Treatise on Man*.[4] Adolphe Quetelet (1796–1874) was a Belgian astronomer and social scientist (in today's terms), who applied the law of errors to human characteristics (height, weight, etc.). Prior to this, the error law or 'normal distribution' had only ever been applied to measurement errors—to apply it to the actual physical things themselves constituted an enormous advance.

Maxwell, of course, understood that the mechanism for the smearing of atomic/molecular speeds arose from the near-infinite number of collisions between molecules. However, he didn't work his way up from the actual collision equations—there were far, far too many of them. Rather, he employed a 'top-down' approach in which he made various assumptions such as that molecules in a certain volume element before collision would have speeds in a certain range, and so on. (By the way, the suggestion that Maxwell could just as well have considered a distribution in atomic mass rather than speed is quite wrong: Maxwell knew it *had* to be speed.[5])

As to the use of probability, Maxwell felt that he had to justify this depravity:

> [probability calculus], of which we usually assume that it refers only to gambling, dicing and betting, and should therefore be wholly immoral, is the only *mathematics for practical people*, which we should be.[6]

As the starting assumptions were randomizing ones (i.e. that the starting speeds were independent of each other,* called the assumption of 'molecular chaos'[7]), then it was not surprising that the normal distribution (a bell-shaped curve) resulted. However, the new statistical or probabilistic methods often led to very counterintuitive outcomes. For example, Maxwell's kinetic theory predicted that the viscosity of a gas was independent of its density. Maxwell was so surprised that, aided by his wife, he performed experiments to check up on this result—it was corroborated.

The Austrian physicist Ludwig Boltzmann (1844–1906) took Maxwell's distribution as his starting point (1872). He looked at the time evolution of a gas and proved that Maxwell's distribution was both stable and unique (i.e. that different starting distributions would all tend towards it and, once arrived at, it would keep reproducing itself). All this occurred via the process of random molecular motions (random in the sense that they were uncoordinated: the bulk motion of, say, a sound wave was ruled out).

Boltzmann took an important step beyond Maxwell—instead of velocity, he considered *energy* distributions and, moreover, energy of a

* But still satisfying the conservation of energy and momentum.

more generalized kind (translations, but also rotations and vibrations, and also potential energy). He had the profound insight that, at equilibrium, the amount of energy tied up in each 'thermal degree of freedom' would be equal: this was the famous 'equipartition theorem'. In this way, the Maxwell distribution for point-particle speeds was generalized to the Maxwell–Boltzmann distribution of molecular energies (in a gas).

Boltzmann was at first (1866) motivated to explain the Second Law and Clausius' entropy principle purely mechanically (i.e. without need of probability theory). He was able to show that introducing an external potential (i.e. doing work on the system) resulted in a reshuffling of the microscopic thermal energies. This was encouraging—Clausius had specifically invented entropy to explain the redistribution of heat consequent upon the performance of work (Chapter 16). However, irreversible processes could still not be dealt with.

For these irreversible processes, Boltzmann ultimately had to turn to Maxwell's probabilistic distribution. In 1872, in his celebrated '*H* theorem', Boltzmann found a functional form, *H*, of Maxwell's distribution. He found that *H* was minimized after any change to the system. Boltzmann didn't hesitate to identify '$-H$' with Clausius' '*S*'—both were maximized. But not everyone was convinced and the question of *S* was still controversial.

Then, in his long paper of 1877, Boltzmann took two more important steps. First, 'as an aid to calculation',[8] he considered the 'fiction'[9] of discrete energy levels, each with the same probability of being occupied. Finally, Boltzmann introduced a radically new approach, although it slipped in without fanfare: he switched from (1) 'the fraction of a time interval during which a given molecule's velocity was within prescribed limits' to (2) 'the fraction of the total number of molecules which had velocities within prescribed limits at a given moment'. In other words, he went from looking at one molecule through time to all the molecules at one time. In this snapshot view of the whole gas (the second approach), he considered the fraction of molecules occupying each energy level.[10]

Boltzmann soon realized that he had to justify the equivalence of these two quite different outlooks. He brought in a hypothesis, which he said was 'not improbable',[11] that, in the course of time, the molecular coordinates and energies would take on all possible values (consistent with the fixed total energy of the gas). (Years later, in 1911, this was called the ergodic hypothesis.[12])

In other words, the solution of the dynamical equations of motion for each of, say, 10^{24}molecules was assumed to be completely equivalent to the non-dynamical question 'How many permutations of molecule energy-assignments will yield a certain energy profile?' The problem of the distribution of energies was thereby reduced to one of combinatorial analysis. The counting of combinations is a standard procedure in statistics and leads to each energy profile having a certain statistical weight, log *W*. This is how Boltzmann's famous equation entered physics:

$$S = k \log W \tag{17.1}$$

where *S* is the entropy, *W* is the statistical weight, *k* is a constant of proportionality (Boltzmann's constant), and 'log' denotes the natural logarithm. This is the equation on Boltzmann's tombstone, but he didn't write it exactly like this and he never had a constant called *k*—that was introduced later by his student, Max Planck.

That this definition of entropy has any overlap with Clausius' macroscopic definition is not obvious. We shall make the best case we can in Chapter 18, 'Difficult Things'. However, Boltzmann himself had no hesitation in identifying his microscopic formulation with Clausius' *S*—both tended towards a maximum value for real processes, and both resulted in a final condition where the energy was dissipated, that is, distributed as uniformly as possible. This was momentous: Boltzmann's statistical approach supplied a mechanism whereby the Second Law of Thermodynamics could be explained.

Physicists of today may reel and gasp in appreciation, but Boltzmann's contemporaries were slower to understand what he had accomplished. First of all, we must remember that molecules had still not been detected. Secondly, that macroscopic irreversibility, in other words a preferred direction in time, could come out of *reversible* molecular collisions, was—and still is—deeply puzzling. This will be discussed further in Chapter 18. For now, we return to energy.

The relevance of Boltzmann's work to the new concept of energy was likewise profound. He recognized that energy was a more telling parameter than velocity. In fact, his equipartition theorem showed that energy is IT, the paramount parameter. At equilibrium, tiny dollops of energy are constantly being redistributed by a hive of random activity (billions of collisions per second for the 'air' molecules in my study, for example), blind as to whether the energy is doled out in rotational, translational, or vibrational form; blind also to different types of gas; blind even to the huge mass of a macroscopic object such as a grain of pollen. The only important thing is that, at equilibrium, the average thermal energy per degree of freedom is a constant and equal to the temperature—in fact, it *is* the temperature.

Boltzmann realized that his statistical approach could apply to more than just gases, but it was an American, Willard Gibbs (1839–1903), who vastly extended it to other thermodynamical systems (heterogeneous mixtures and chemical, elastic, surface, electromagnetic, and electrochemical phenomena). Gibbs is therefore rightly called the founder of statistical thermodynamics and of physical chemistry. In his monumental work of 1878, *On the Equilibrium of Heterogeneous Substances*,[13] Gibbs applied two techniques from earlier eras: the idea of phase space (growing out of Hamilton's (p,q) space) and the variational calculus (as, at equilibrium, infinitesimal variations in entropy or energy had to sum to zero—reminiscent of Johann Bernoulli's principle of virtual work from 150 years earlier).

Electromagnetism

In addition to the First and Second Laws of Thermodynamics, the other major achievement of nineteenth-century physics was Maxwell's theory of electromagnetism, developed from 1865 to 1873. Maxwell tried to incorporate Faraday's 'field' lines and other experimental results in a mathematical way. He introduced a new feature, the displacement current, which extended itself beyond the confines of the apparatus (conductors and magnets etc.) and into the surrounding empty space. It was as if there were stresses and strains within space itself. Moreover, this space *did* appear to contain energy, the energy of the electromagnetic field. As Maxwell wrote:

> In speaking of the Energy of the field, however, I wish to be understood literally…The only question is, Where does it reside? On the old theories it resides in the electrified bodies, conducting circuits, and magnets, in the form of an unknown quality called potential energy, or the power of producing certain effects at a distance. On our theory it resides in the electromagnetic field, in the space surrounding the electrified and magnetic bodies, as well as in those bodies themselves.[14]

Although there is now no doubt that energy exists in the field—for example, things warm up in the sunshine—there is still a problem with exactly where the energy resides. For example, there is the paradox that the energy associated with a point charge is infinite.[15] Also, in certain experimental configurations (e.g. a bar magnet with an electric charge nearby), the energy must perpetually circulate in order for angular momentum to be conserved.

Furthermore, it turns out that in order to make Maxwell's theory of electromagnetism consistent with Einstein's Theory of Special Relativity (see below), the conservation of energy must be replaced by the more stringent requirement that energy is conserved *locally*. The electromagnetic energy in a given small volume (the energy density) must then be balanced against the energy flowing into or out of that volume (the flux of energy). This energy flux was defined by John Poynting (1852–1914) in 1884 and used to explain the transfer of electromagnetic energy from one place to another through space.

Radiant Energy

Problems with the classical theory of heat occurred at both high frequencies (the so-called ultraviolet catastrophe) and low temperatures (the anomalous specific heats). The first problem concerned the radiation from black bodies and was resolved by Boltzmann's student, Max Planck (1858–1947), in 1900. He was forced to adopt a radical heuristic—the energy of radiation must come in discrete chunks or quanta. It was Boltzmann's earlier quantization of the energy distribution in a gas that gave Planck the permission to take this unpalatable step—a bit like allowing running only at certain speeds, or stair treads only at certain heights.

Planck's new theory had the radiation energy given by

$$E = hv \tag{17.2}$$

(v is the frequency of the radiation and h is a constant, now known as Planck's constant). This corrected the anomalies, but it is remarkable to think that it made any difference at all, and to gross phenomena, considering that the spacing of the energy intervals was tiny (h is around 6.6 $\times 10^{-34}$ J s^{-1}). Planck had, of course, started off the quantum revolution. 'Small', 'smaller', and 'smallest' are all relative terms, but when the relevant physical quantities are of the order of h, you know that this is *absolutely* small, the domain of a new mechanics called quantum mechanics.

Spurred by the phenomenon of photoelectricity, Albert Einstein, in 1905, postulated that Planck's equation applied not only to emitted radiation but also to absorbed and transported radiation—in other words, the radiation, or light itself, was quantized, that is, particulate. (The light particle was later called the photon.) If light waves could be considered as particles, then maybe other particles could be considered as waves? This was suggested by Louis de Broglie in 1924, and experimentally confirmed for electrons by Davisson and Germer in 1927.

The trouble with the specific heats was also resolved by the newly emerging quantum theory. Dulong and Petit had found that the specific heat per atom was a constant for different solids (around 25 J per atom per degree; see Chapter 10). However, for diamond and some other hard solids the specific heat was found to decrease with temperature rather than remain constant. Einstein resolved this in 1907 by adopting Planck's quanta of energy again. Quantization also explained other anomalies, such as the discontinuous decrease in the specific heat of liquified gases: at very low temperatures, certain energy levels were 'frozen out'—the quantum of energy simply wasn't large enough for even low-lying levels to be reached. Finally, polyatomic gases were found to have a larger specific heat capacity than monatomic gases—due to the greater number of thermal degrees of freedom in the former case. So, as commented upon in Chapter 10, the specific heat was proving to be a macroscopic window into microscopic structure.

The Quantum World

A new feature of the quantum-mechanical realm, with ramifications for energy, was Heisenberg's Uncertainty Principle (1927). According to Werner Heisenberg (1901–76), the energy and the time could only be jointly determined to precision $\hbar$:

$$\Delta E \times \Delta t \geq \hbar/2 \tag{17.3}$$

where $\hbar = h/2\pi$. Energy conservation could thus be contravened to the extent ΔE as long as this loan was paid back within time Δt. For example, one joule could be 'borrowed' for around 10^{-34} s.

Erwin Schrödinger (1887–1961) saw that Hamilton's optico-mechanical analogy would be ideally suited to coping with de Broglie's wave-particle duality. He therefore adapted Hamilton's mechanics for use in the quantum world and so derived the famous Schrödinger wave equation. Hamilton's multiplicity of possible paths had mapped out the gross features of phase space (see Chapter 13). Likewise, Schrödinger's wave mechanics yielded the overall conservation and symmetry relations for the given problem. This approach was to be especially useful in the quantum-mechanical domain where almost all the problems are in practice too hard to solve exactly. The best line of attack is to delimit the problem and make sure that the relevant conservation and symmetry relations are satisfied, and then use some appropriate method of approximation. The most important conservation principle is still the conservation of energy, and so the 'Hamiltonian' is as crucial in quantum mechanics as it was in classical mechanics. (One earlier wave theory—due to 'BKS'—had to be discarded, in part because energy was only conserved on average.)

There was one major difference between the classical and quantum worlds, however. In the latter case, Hamilton's 'possibilities' were reinterpreted as 'probabilities' for the wave describing the whereabouts and momentum of just one quantum-mechanical particle. This interpretation was introduced by Max Born (1882–1970) in 1926, but Schrödinger was never happy with it.

Richard Feynman (in 1948)[16] developed Hamilton's methods even further and adopted a fully variational approach: he calculated the action—the integral over the entire path in time (the 'history')—for all possible paths/'histories' of the quantum-mechanical particle.

Special and General Relativity

The laws of thermodynamics were very influential in all of Einstein's thinking. He wrote, 'classical thermodynamics . . . is the only physical theory of universal content concerning which I am convinced that, within the framework of applicability of its basic concepts, it will never be overthrown'.[17] In the same way as thermodynamics was a 'theory of principle',[18] so were Einstein's Theories of Special Relativity (SR) and General Relativity (GR).

In SR, the fundamental laws of physics had to be cast in a way that ensured that different, valid, viewpoints (frames of reference) were equivalent. This was Einstein's Principle of Relativity. There was the additional principle that the speed of light was a constant for all such observers.

What are the consequences of SR for energy?

In SR, the concepts of absolute simultaneity (of spatially separated events) and of absolute and independent space and time all have to be sacrificed. These shifts have dynamic implications. Not only must energy be conserved locally (as mentioned above), but energy and momentum must be conserved together. In the same way as space and time have no separate existence but form a new continuum (space–time), so energy and momentum are melded together and, as so often in physics, this heralds a new mathematical 'object', the energy–momentum tensor. The eighteenth-century fight for supremacy, lasting almost 100 years, between Newton's momentum and Leibniz's 'energy' is finally resolved—both are important.

We have saved the most famous until last. A requirement of making Maxwell's theory of electromagnetism consistent with SR is that the speed of light, c, should never be surpassed. In effect, bodies become more massive as they get faster (i.e. as their kinetic energy increases) and at just such a rate that c can never be reached. In other words, there is a new feature, a *mass-dependence of kinetic energy*. The energy, E, must increase in proportion to the moving mass, m. Everyone knows that the constant of proportionality is c^2:

$$E = mc^2 \tag{17.4}$$

Einstein considered that this was the most profound result to come out of SR.[19] No longer was there Kant's divide between 'stuff' and 'activity'. There had been earlier suspicions of this energy–mass equivalence (Thomson, Henri Poincaré, and Friedrich Hasenöhrl), but Einstein was the first to boldly postulate that it was universal, applying to all types of energy and all types of mass. This was aesthetically more beautiful and also was suggested by the (by now well-accepted) interconvertibility between all the different forms of energy.

One consequence of this universality of $E = mc^2$ was that, as Einstein observed, 'the conservation of mass is now a special case of the conservation of energy'.[20] Prior to Einstein, there was no logical need for inertial mass to be conserved (inertial mass—the 'm' that occurs in $F = ma$—is a measure of a body's resistance to motion; see Chapter 7). Conservation was just taken for granted because of Newton's tacit assumption that inertial mass was the same thing as 'matter', which was itself assumed to be made from indestructible components. Now, with Einstein's equation, the conservation of inertial mass was made intuitive—it was just part of the conservation of energy.

If the kinetic energy (and its associated mass-equivalent) is reduced to zero (i.e. the body is in its own 'rest frame'), then the stationary body still has some mass 'left over'. This, the rest mass, m_0, is the usual mass with which we are all familiar, the mass that may be weighed on kitchen scales and so on. Einstein postulated that as well as leftover mass there was also leftover energy, even when the *kinetic* energy has gone to zero. To put it another way, he postulated that the rest mass could be converted into energy (the 'rest energy', m_0c^2). This was not required by SR—for example, the rest mass might have been indestructible. (Later, Feynman showed that the annihilation of matter and antimatter had to be possible and so rest mass was not indestructible.[21])

It has always seemed mysterious, in classical physics, how the kinetic energy can change in value from one frame of reference to another: now, for the first time, there is a natural zero point to the kinetic energy, the rest energy, m_0c^2. What is still curious, however, is that this zero-point energy is vastly larger than the commonplace kinetic energy. (For example, the rest energy in a gram of matter is about the same as that released in the bomb at Hiroshima.) How is it that we are normally (thankfully) completely oblivious to this vast zero-point energy? It is simply due to the fact that this energy is, for the most part, remarkably stable. Perhaps this is no more surprising than the fact that electrical effects were virtually unknown about for millennia, even while the electric force is 10^{40} times stronger than the gravitational force.

We have seen how Einstein postulated that both inertial (moving) mass and rest mass had energy equivalents. Going the other way, all types of energy should have a *mass* equivalent. For example, light and such forms of 'stationary' energy as heat and chemical bonds should all be massy.

With Einstein's Theory of General Relativity (GR), there came both experimental and theoretical confirmation of this. For example, experimentally, light rays were found to be bent by the Sun's gravity, showing that light does have weight. Also clocks—for example, atomic clocks—were found to run slower when near a large centre of gravity, as if the clock 'spring' was heavier.

Theoretically, anything weighty should be affected by gravity (that much was still true). But…everything *is* affected by gravity. Therefore everything, even inertial mass, does have weight. This last seems tautologous, but it is only another tacit assumption of Newton's that makes it seem so. For Newton, the identicality of the mass in $F = ma$ and the masses in his Law of Universal Gravitation was a strange coincidence. The fact that all objects, regardless of their mass or their chemical composition (feathers, cannon balls, etc.), all fall towards Earth at the same rate was part of this coincidence. Now, with GR, all is explained. All these masses—whether inertial, gravitational, or rest mass—all move in exactly the same way because their motion is determined by *geometry*. They are all following the same 'railway track', all subject to the same curvature of space–time.

In going from Newtonian mechanics to SR, we have found that energy and stuff are really the same sort of thing; now, in going from SR to GR, we find that energy, stuff, *and space–time* are all interrelated. Moreover, mass-energy is not only subject to geometry but may be the very cause of it (i.e. the cause of curvature in space–time).

However, there is still a paradox as regards the exact location of this gravitational mass-energy, this energy of the field. Electromagnetic energy explodes to infinity when confined to a point charge. On the other hand, gravitational field energy disappears altogether as soon as one looks too closely—as everywhere and everywhen space–time is *locally* flat (ignoring black holes and the Big Bang).

Finally, on current theories (2010), there is another location problem—apparently, around 95% of the energy is missing. Everything we can see and observe on Earth and in space makes up only 5% of the total. The remaining 95% is dark matter and dark energy, and is still a mystery.

Interactions between the Physicists in the Second Half of the Nineteenth Century

In 1880 a young American physicist, Henry Rowland, measured the mechanical equivalent of heat using the paddle-wheel method, and obtained a value slightly higher than Joule's value. Joule wrote to Thomson:

> I don't think my result can be wrong... I have received his [Rowland's] thermometer and shall as soon as possible compare it with mine... the fineness of the markings is much inferior to mine.[22]

Rowland had carried out Joule's paddle-wheel method over a range of temperatures up to 37°C. The mechanical equivalent fell, but this was actually due to the fact that the specific heat capacity of water was not a constant over this temperature range. In other words, Joule's method could be used to determine the specific heat of water *assuming* energy and its conversion and conservation. Physics had come full circle.

For example, Poynting took energy for granted when he stated that the method of falling weights (Joule's paddle-wheel experiment) was the best way of determining specific heats, as there was nothing hypothetical about it.[23] There *was* still a hidden assumption, though—was the strength of gravity, for example between Manchester and Baltimore, a constant? Joule suggested repeating the experiment at Greenwich:

> the latitude of Greenwich is impartial in just skirting the south of Ireland and the north of France and also the north of Austria; or at least it used to do so before the Devil with his wars displaced the countries and I don't buy fresh maps every year.[24]

Despite the political correctness of Greenwich, the point is that no *absolute* determination can ever be found without taking something as a standard. The French mathematician and philosopher Henri Poincaré (1854–1912) took this to its logical conclusion (see Chapter 18). He saw that the very generality of the principle of energy conservation meant that it had to be assumed and could therefore never be confirmed by experiment.

The three Scots, Thomson, Tait, and Maxwell, formed a cosy trio, communicating by postcard and referring to each other by the nicknames T, T′, and dP/dt, respectively. Thomson and Tait were the spearheads of a crusade—to bring the new doctrine of energy and energy conservation into natural philosophy. They wrote a textbook, *Treatise on Natural Philosophy*,[25] published in 1867 and commonly referred to as T& T′, which set the programme for classical physics in the nineteenth century. However, Tait was shamefully keen to appropriate the whole of physics to British and, above all, Newtonian origins. There is a risible episode in which he searched through the whole of Newton's *Principia* saying that it (energy) had got to be in there somewhere.[26] ('It' wasn't: as we have shown in Chapters 3 and 7, Newton missed discovering energy.)

Tait totally misunderstood Clausius' concept of entropy, calling it the 'available energy'. Thomson was conspicuously silent. Maxwell relied on Tait to keep him informed scientifically and so managed to write a whole book on heat (*The Theory of Heat*[27]) while being mistaken about entropy and not mentioning Clausius as its discoverer. Clausius protested, but it wasn't until Gibbs took up Clausius' cause that anything happened. Maxwell 'was led to recant'[28] and to acknowledge Clausius' priority and his achievements (which Maxwell was very happy to do—he wrote, 'I mean to take such draughts of Clausiustical Ergon [work] as to place me in . . . [a] state of disgregration'[29]). John Tyndall, translator of Clausius' works into English, was another physicist who explained to T, T′, and dP/dt that Clausius was a genuine and good fellow. (Tyndall had also been active in promoting Mayer's cause; see Chapter 14.)

Maxwell and Boltzmann's work on the kinetic theory led to the Maxwell–Boltzmann distribution, but they never met or even corresponded. Maxwell respected Boltzmann's ideas, but found his work easier to follow by *not* reading Boltzmann's long papers, confiding to Tait '[I could] put the whole business in about six lines'.[30] Boltzmann, meanwhile, revered Maxwell and said of Maxwell's (electromagnetic) equations, 'Was it a God who wrote these signs?',[31] a quote from Goethe's *Faust*.

Maxwell was 'absurdly and infuriatingly modest'.[32] In a survey talk in 1870, he praised Thomson's vortex theory of electricity at great length and then only briefly mentioned his own work, saying 'Another theory of electricity which I prefer . . .'[33] (Maxwell's theory of electromagnetism is now reckoned to stand alongside the work of Newton and Einstein as one of the most outstanding achievements in physics.)

In the nineteenth century the mechanical world-view, introduced by Descartes in the seventeenth century (see Chapter 3), was struggling to survive. Maxwell's theory of electromagnetism seemed to need an ether in which the electromagnetic waves could travel, but the mechanical properties required of this ether were impossible. Another new feature was the arrival of scientific laws with a cosmic direction given to time. First, there was Fourier's law of heat conduction, then the Second Law of Thermodynamics, and finally, the biggest revolution outside the physical sciences, Darwin's theory of evolution in 1859.

It is hard now to appreciate the novelty of this cosmic time-dependence. Laplace and Lagrange had carried out calculations confirming that the planetary orbits were stable; the uniformitarian geologist Charles Lyell (1797–1875) did not consider that a hot Earth core contradicted his assumption of geological uniformity over very long, possibly infinite, timescales; and Clausius wrote in 1868, 'One hears it often said that . . . the world may go on in the same way for ever.'[34]

Newton, as so often, was a rare exception, a solitary giant and less dogmatic than the 'Newtonian world-view' that succeeded him. He *had* noticed that 'motion is always on the decay' (*Opticks*, Query 31) and was unable to resolve this except by suggesting that a Creator might patch things up. Leibniz was scornful of Newton's need for a 'divine watchmaker' to wind up his clockwork universe from time to time.

Thomson was the first to fully embrace the possibility of an overall direction for time, presumably preferring it to either Laplacian determinism or blind chance—both were antithetical to religious sensibilities, not allowing room for a 'controlling intelligence'.[35]

In 1851, quoting from the Bible, Thomson wrote: 'Everything in the material world is progressive... "The earth shall wax old &c".'[36] The word 'progressive' is optimistic, but Thomson followed this up with his landmark synopsis of the Second Law in 1852: '[there is] a universal tendency to dissipation' and the Earth will be, in some future era, 'unfit for the habitation of man' (see Chapter 12). Helmholtz was one of the very few who understood its import: 'we must admire the sagacity of Thomson, who... [using but a] little mathematical formula which speaks only of the heat, volume and pressure of bodies, was able to discern consequences which threatened the universe... with eternal death'.[37]

There were some physicists who refused to countenance this bleakness. Rankine suggested that perhaps radiant energy could be reconcentrated at the boundary of the universe and then reflected back to us; and Loschmidt (see below) was another who objected to the 'terroristic nimbus'[38] of the Second Law. But there were no public outbreaks of despondency and despair—in fact, there was almost complete silence regarding the Second Law. The ideas were simply too big to be understood in their own time. It wasn't until the popularizations of James Jeans and of Arthur Eddington in the twentieth century that the implications of the Second Law of Thermodynamics entered the public domain.

There were a few nineteenth-century philosophers who paid attention to physics: one was Friedrich Nietzsche (1844–1900), who formulated his theory of eternal recurrence, and another was Herbert Spencer (1820–1903). The latter incorporated the First and Second Laws of Thermodynamics into his ideas of evolution towards 'the greatest perfection and the most complete happiness' and 'alternate eras of Evolution and Dissolution'.[39] Maxwell didn't comment on these particular conclusions, but thought that it was useful to have the views of non-scientists: 'Mathematicians, by guiding their thoughts always along the same tracks, have converted the field of thought into a kind of railway system, and are apt to neglect cross-country speculations.'[40]

Thomson had an input into the other two 'progressive' laws of cooling and of evolution. By using Fourier's laws of heat conduction and radiative cooling, Thomson showed that the Earth had to be less than 200 million years old—not old enough to allow for Darwinian evolution. Thomson undoubtedly had an agenda, which was to discredit Lyell's and Darwin's theories. (Fourier, in 1820, had made a similar estimate, but thought it was absurd.[41])

Thomson's calculations were largely correct (given that he had to make a number of starting assumptions) but, of course, he hadn't known about radioactivity. He lived long enough to learn that radioactivity, discovered just before 1900, acted as another source of terrestrial heat. This extended the age of the Earth sufficiently to allow for Darwinian evolution. Thomson (now Lord Kelvin) was not amused and, apparently, never fully accepted that his 'most important'[42] work had been overturned.

On the other hand, substituting a negative time into Fourier's equations showed that the Earth must have cooled down from a special initial state. Thomson thought that this amounted to a mathematical proof for a 'free will having power to interfere with the laws of dead matter'.[43]

The biggest paradox in thermodynamics was that concerning reversibility. The Second Law indicated an absolute direction in time, and yet it derived from the interactions of microscopic particles governed by laws that were *independent* of a direction in time. In other words, macroscopic irreversibility appeared to arise out of microscopic reversibility.

The Scottish trio were the first to discuss these matters. Maxwell wrote to T and T′ (in 1870):

> if you accurately reverse the motion of every particle . . . then all things will happen backwards . . . the raindrops will collect themselves from the ground and fly up to the clouds, &c &c and men will see all their friends passing from the grave to the cradle till we ourselves become the reverse of born, whatever that is.[44]

Maxwell fully resolved the paradox to his own satisfaction: he saw that while the microscopic laws of collision were completely reversible, his problem of the distribution of velocities hadn't been solved by these laws alone. In dealing with the near-to-infinite numbers of molecules, he had had to make various assumptions about the starting velocity and position distributions. He realized that it was in these very assumptions that a macroscopic arrow of time had slipped in. (As a 'thought experiment', Maxwell invented a fictitious microscopic creature, his 'demon', specifically to demonstrate that one couldn't cheat the system and the Second Law *had* to be statistical in nature.)

Maxwell ruefully watched the German school (Clausius and others) in their attempts to find a purely deterministic explanation of the Second Law, writing to Tait: 'It is rare sport to see . . . the German Icari flap their waxen wings in nephelococcygia [cloud-cuckoo-land].'[45]

Clausius never accepted Boltzmann's statistical interpretation of entropy. Sadly, it was not appreciated in Boltzmann's lifetime (the first explanatory paper, by Paul and Tatyana Ehrenfest, appeared only in 1911 and the tombstone bearing Boltzmann's famous equation was erected in the 1930s). The barriers to understanding were manifold. As well as the reversibility conundrum, thinking statistically was still alien and the results not at all intuitive. There was, in addition, the problem that the German school of energeticists and positivists didn't even believe in atoms ('I don't believe that atoms exist!'[46] said Mach at a meeting in Vienna in 1897.)

Boltzmann received friendly criticism in Britain, but in Germany he felt he was either misunderstood or ignored. To the great Dutch physicist, Hendrik Lorentz, he wrote (in 1887): 'I am very happy to have found in you a person who works to develop my ideas on the kinetic theory of gases. In Germany there is nobody who correctly understands these things.'[47]

Boltzmann did receive one very penetrating criticism (in 1876) from an almost-German, the physicist Josef Loschmidt (from Bohemia, then part of the Austro-Hungarian Empire). (Loschmidt was most certainly not an anti-atomist, having, in fact, determined Avogadro's number; see Chapter 10.) Loschmidt puzzled over the same reversibility paradox as the Scots; namely, what happens when the motion of every atom is exactly reversed? In fact, it was in pondering this question that Boltzmann was led to try out his radically new combinatorial approach.

The German mathematician Ernst Zermelo (1871–1953) raised a similar paradox in 1896. Poincaré had shown with his 'recurrence theorem' that any mechanical system will eventually return to its initial state. Zermelo saw that this could lead to a reduction in entropy, if you waited long enough. Boltzmann countered 'you should wait that long'[48] (over $10^{10 \text{ billion}}$ years for a gas sample to spontaneously divide into its nitrogen and oxygen components).

Boltzmann saw that both Loschmidt and Zermelo could be answered in the same way: reversed, entropy-decreasing, motions *can* occur, but they are extremely improbable. Planck, while vehemently opposed to the energeticists, couldn't accept Boltzmann's statistical solution, as it implied that the Second Law wasn't *absolutely* true (i.e. true in every single case).

Another German was Hermann von Helmholtz, whom Boltzmann admired but found imperious and formal. (When dining at the Helmholtz residence, Boltzmann apparently picked up the wrong piece of cutlery and Frau von Helmholtz said, 'Herr Boltzmann, in Berlin you will not fit in.'[49])

Outside of the physical sciences, Darwin's theory of evolution was, as we have said, the greatest scientific advance of the nineteenth century. Boltzmann was a great admirer of this theory, perhaps seeing in it another example of a theory in which an innumerable succession of random events could lead to an irreversible outcome. He wisely observed: 'The overall struggle for existence of living beings is . . . a struggle for entropy [rather than for energy or raw materials]'.[50]

Willard Gibbs was an exceedingly quiet American, a professor at Yale, unmarried, and living his entire life (apart from foreign travels) in his childhood home, less than a block away from the university. His 'Gibbs paradox' convinced him, in agreement with Maxwell and Boltzmann, that 'the impossibility of an uncompensated decrease of entropy seems to be reduced to improbability'.[51] He was exceedingly modest as well as quiet, claiming that 'anyone with the same desires could have made the same researches'.[52] With this combination of modesty and 'extreme economy (one might almost say parsimony) in the use of words',[53] it is hardly surprising that Gibbs appealed to Maxwell. Maxwell made a clay and plaster cast of Gibbs' three-dimensional thermodynamic surface for water and sent it to him in New Haven, in 1874. (It can still be seen, in a display case at Yale University.)

It is hardly surprising that Boltzmann developed a strong antipathy towards philosophy. This is captured in his provisional title for a talk: 'Proof that Schopenhauer is a stupid, ignorant philosopher, scribbling nonsense and dispensing verbiage that fundamentally and forever rots people's brains'.[54] (This was actually a paraphrase of Schopenhauer himself on Hegel.) Boltzmann also parried a criticism about his lack of mathematical elegance, saying that elegance was for tailors.

Maxwell was one who could perhaps have 'saved' Boltzmann (his last researches were on Boltzmann's work) but he died young, at age 48, at exactly the same age and of the same condition (stomach cancer) as his mother. In 1906, Boltzmann committed suicide while on holiday with his family. While the rejection of his life's work must have played some part in this, the major cause was the fact that Boltzmann suffered from some form of manic depression ('neurasthenia' to the Victorians). Only a few weeks after Boltzmann's death, there was unequivocal proof, in

Einstein's and in Marian Smoluchowski's analysis of Brownian motion, that atoms do exist.

Public Energy

Apart from the physicist's conception of energy, energy has arrived in another way. We now have industrialized economies and power supply networks; 'energy' has come into the public domain and everyone understands what it is. Did any of the nineteenth-century discoverers of energy anticipate this?

When Faraday was asked by the Prime Minister, William Gladstone, what was the use of electricity, Faraday famously replied that one day the government would tax it. On the other hand, Joule's researches seemed to indicate that electric power would never compare to steam power as regards efficiency.

Electric telegraphy began around mid-century and William Thomson, hired by the telegraphy company, was instrumental in the cross-Channel and cross-Atlantic lines. However, these early lines mostly connected public buildings or central post offices. The idea of a communications network, requiring *electricity*, and serving almost everyone on the planet, was inconceivable (there is also the Catch-22 that a partial network isn't very useful). In a similar vein, the Victorian science fiction writer, H.G. Wells, correctly predicted that an aircraft was technically feasible but he also predicted that it would never be economically viable—the quantity of oil required was simply absurd and the idea of a global politico-economic nexus arising specifically to extract this sticky brown stuff—fantastical.[55]

The economist Stanley Jevons wrote in 1865 that the rate of using coal had increased by 3.5% every year, and if the trend continued then Britain would run out of coal[56] by 1965. Osbert Reynolds, Joule's biographer, wrote that coal could always be imported (the problem of dwindling *global* supplies never occurred to him).

In general, Reynolds saw the energy problem as one of losses during transmission rather than of supplies at source. He compared transmission methods such as belting, shafting, compressed air, or hydraulic mains.

When it came to the transmission of electricity along wires, Reynolds was dismissive: 'electric transmission is far inferior to the flying rope'.[57] (The modern mind boggles at the thought of the countryside criss-crossed by miles of whizzing rope.)

Eventually, it was understood how losses during transmission could be overcome. Joule had shown that heating (the main loss) was proportional to I^2R; so if the current was reduced in transit, the losses would be lessened.[58] A lower current could be achieved by 'stepping up' the voltage, easily done by a transformer (invented in 1883), which in turn required alternating current (the AC alternator was invented by Nikola Tesla in 1888). All this came to pass—but what was the need for *electric* power as opposed to, say, piped coal-gas or deliveries of solid fuel?

The turning point was the discovery of the electric light bulb (independently by Joseph Swan and by Thomas Edison in 1879). The capital outlay per household was not enormous and the light was easier to 'switch on', safer, cleaner, and above all brighter than the pre-existing gas lamps. Once electricity was supplied to each home, then the market for electric appliances was opened up. The electricity was mostly generated in coal-fired power stations employing a very large steam engine. Even today, the steam engine, in the form of a steam turbine, is at the heart of the coal-fired power station. Thus the engines of both Watt and Joule still live on.

18

Impossible Things, Difficult Things

Why, sometimes I've believed as many as six impossible things before breakfast.

White Queen, in *Alice's Adventures in Wonderland*, by Lewis Carroll

Impossible Things

From its history, we have found that the concept of energy has not been easily won. But physics is not the same thing as history of physics: isn't it sufficient to learn about energy from a physics textbook? Sufficient—maybe; easy—no. A famous physicist, I think it was George Gamow, once said that to try to understand thermodynamics he went through it three times: the first time, he understood it but thought he'd better have another go, just to be sure; the second time, he realized he didn't understand it; the third time, he realized he never would understand it. In mildly humorous vein, we look at two classic texts: *The Elements of Classical Thermodynamics*[1] by Brian Pippard (1920–2008) and *Heat and Thermodynamics*[2] by Mark Zemansky (1900–c.1985).

Zemansky defines the system as 'a restricted region of space or a finite portion of matter',[3] which may be separated from the 'surroundings' by walls that are diathermal or adiabatic (they respectively do or don't let heat through).

Remembering all the difficulties, first in the understanding of heat and then in the discovering of energy, it is a little disconcerting to come across the reverse viewpoint wherein energy is *assumed* and heat merely *defined* as the difference between internal energy and work, W:

[heat is] equal to the difference between the internal-energy change and the work done.[4]

Now, what is work? Work is anything that causes external changes in macroscopic quantities such as are ultimately equivalent to the raising or lowering of weights, the winding or unwinding of a spring, or, in general, the alteration of the configuration of some external mechanical device.[5] It is curious that something so anthropic and vague as 'a device' comes into physics in this way.

Whenever the internal energy, U, of a system is changed by pure work, W, then we must have $\Delta U = \Delta W$, and this is true by whatever route the change has been accomplished.* In fact, this amounts to nothing less than the First Law of Thermodynamics. However, it is surprising to find it admitted that 'Unfortunately, it does not seem that [careful] experiments of this kind have ever been carried out'[6] and 'accurate measurements of adiabatic work along different paths between the same two states have never been made'.[7]

As heat is defined as what makes up for the discrepancy between ΔU and ΔW, it is crucial to make a clear distinction between heat and work. However, this distinction can depend on the choice of system boundary. In Fig.18.1(a), the system has had work done on it (from a falling weight), whereas in Fig. 18.1(b) the system has been heated.

We have had to assume that the pulleys are frictionless and that the generator and connecting wires have no resistance. It seems that work and electricity can pass through adiabatic walls, and also a falling weight presupposes a gravitational field—that crosses walls.

How can we be assured that no heat has been added to a system (and thus check up on the First Law)? Answer: by enclosing the system within adiabatic walls. Yes, but how do we know that such walls are truly adiabatic? Because they don't let any heat through...

Diathermal walls *do* let heat through and this is how equilibrium is attained between adjoining systems. But if there are no changes in either system, is this because the walls are not diathermal or because the two systems are already in equilibrium?

What is equilibrium? Equilibrium is the state that has been achieved when there is no further change in the macroscopic properties of a given system or of any number of systems connected by diathermal walls. But how long do we wait? And if there *is* a change, then is it a departure from equilibrium or just a fluctuation?

* 'Δx' means 'an increment in x'.

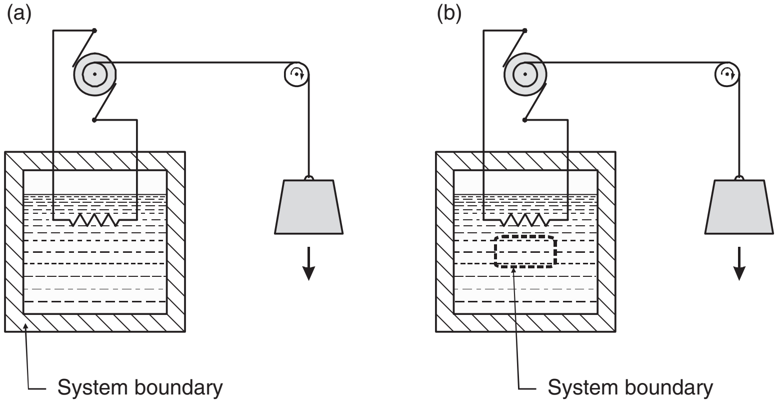

Fig. 18.1 Heat and work depend on the choice of system boundary (redrawn and adapted from Zemansky, *Heat and Thermodynamics*, 5th edn).

From equilibrium, we come naturally to temperature. Brian Pippard introduces temperature without any regard to heat. All we need to know is: what equilibrium is, that there are such things as diathermal walls, and that the Zeroth Law of Thermodynamics applies:

The Zeroth Law of Thermodynamics
If, of three bodies, A, B and C, A and B are separately in equilibrium with C, then A and B are in equilibrium with one other.[8]

Consider, as a test 'body', the usual case of a gas characterized by its pressure, P, and volume, V. Bring this test body into diathermal contact with the 'body' to be investigated—say, the air in the lounge—and adjust the P and V values until equilibrium has been achieved. Then, systematically vary the volume of the test body and for each V adjust the P until equilibrium has again been achieved.

It will be found that all the resulting (P,V) pairs lie on a curve for which some function, $\phi(P,V)$, has a constant value (Fig. 18.2(a)).

Now choose some other 'body' to investigate—say, 'water in the bath'—and repeat the process. The new (P,V) pairs lie on another 'curve of sameness' (Fig. 18.2(b)). Repeating the process will lead to a whole family of such 'curves of sameness.' Finally, label each isocurve with a number. *This number is the temperature*. The number can be 'chosen at will…provided there is some system, however arbitrary, in the labelling'.[9] However, once the labelling has been carried out, then it is fixed

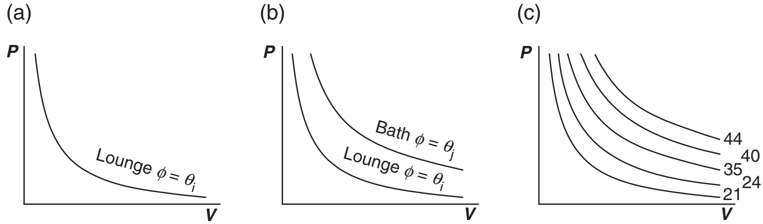

Fig. 18.2 'Curves of sameness': (a) one isocurve; (b) two isocurves; (c) a family of isocurves (after Pippard in *The Elements of Classical Thermodynamics*).

even if the test body is replaced by a different sort of test body, such as a mercury-in-glass thermometer or a platinum resistor.

So far, we have not touched on any link between temperature and heat. In order to do this, Pippard brings in an extra premise: when two bodies are in diathermal contact they come into equilibrium by the 'transfer of heat' from one to the other. As we don't know what heat actually is, then we can't be sure which way it is being transferred. However, once we decide that it is transferred *from* hot to cold, then subsequent experiments show that it is always transferred in this direction. It then only remains to set up a scale of hotness such that bodies that are hotter are also at a higher temperature. We won't elaborate but, in order to prove that setting up such a scale is possible, Pippard uses a *reductio ad absurdum* argument along with the converse of the Zeroth Law. One has the feeling that some arcane game of logic is being played out.

In order to establish the Second Law of Thermodynamics, an impossible scenario must be imagined—the Carnot cycle or ideal heat-engine (see Chapter 12). The gas must be ideal and the piston must fit perfectly, and move without friction and infinitely slowly (so as not to cause turbulence or discontinuities in the gas density). The pressure outside the cylinder must therefore at all times be only infinitesimally different from the pressure within. The cylinder walls and piston must be perfectly insulating, except when thermal contact with a heat reservoir is required. The reservoirs themselves must be infinite (i.e. unaffected by the heat added to or taken away from them) and only infinitesimally higher or lower in temperature than the gas-engine. When all these conditions are satisfied, the changes are said to be quasi-static.

When we come to the kinetic theory (say, of gases), we have some further idealizations to make. The volume-element under consideration must be small when compared to the total volume but large enough to include a very large number of microscopic components (molecules, atoms, ions, or whatever). At usual temperatures and pressures, a millionth of a cubic centimetre of air contains around 10^{13} molecules—and this is deemed a satisfactory size for a volume element of gas.[10]

There are also assumptions to do with the ideal nature of the gas: the molecules are taken to be small hard spheres that are in perpetual random motion and exert no forces on each other except at the instant of collision. Between collisions, they therefore move with uniform rectilinear motion. The molecular diameter is assumed small compared to the average separation of neighbours (typically, molecules are a few ångstroms across (1 Å = 10^{-8}cm) and their average separation is 50 times this). Collisions are assumed to be perfectly elastic and the container walls perfectly smooth. Actual gases most closely approach this ideal in the limit as the pressure is reduced to zero. (In other words, when there is no gas whatsoever, then it is ideal.)

In the early kinetic theories (such as that of Daniel Bernoulli; see Chapter 7) the molecules were considered as point particles and therefore never collided: equilibrium was slow in coming (it could be arrived at only by collisions with the walls). An even worse problem arises in the case of a cavity consisting of perfectly internally reflecting walls and containing only radiation. The radiation not only doesn't interact with itself but doesn't interact with the walls either. The only way in which equilibrium can be achieved is to (hypothetically) introduce 'an extremely small piece of matter such as a grain of coal'.[11]

It is one thing whether a thought experiment is difficult to carry out, but can it be allowed that it is impossible, even in principle?[12]

Difficult Things

We move from the impossible to the merely difficult and consider a miscellany of perennially difficult topics. This is not a textbook and I shall aim to demystify and explain in woolly words and analogies rather than with mathematics. In so doing, I shall be sticking my neck out from time to time. To quote from Feynman: 'A physical understanding is a completely unmathematical, imprecise, and inexact thing, but absolutely necessary [for a physicist].'[13]

The Laws of Thermodynamics—Are They Empirical?

Let's begin by explaining that the First Law of Thermodynamics and the Law of Conservation of Energy are really the same thing, except that the former applies specifically to heat and work and the latter to all forms of energy.

Both the First and Second Laws of Thermodynamics started from the common experience that perpetual motion of the first and second kinds is impossible (perpetual motion of the first kind refers to an isolated engine perpetually generating work; perpetual motion of the second kind refers to heat flowing from cold to hot unaided). It is because the laws started from such common knowledge that they are said to be empirically based. It was, in fact, the extraordinary broadness of this empirical base that led to the extraordinary broadness of the reach of these laws.

However, the early investigators, in particular Mayer and Helmholtz, considered that the conservation of energy was not just confirmed by experiment but *had* to be true because of the deeper law of 'cause equals effect'. Planck made a similar point:

> the impossibility of perpetual motion of the first kind is certainly the most direct of the general proofs of the principle of energy. Nevertheless, hardly anyone would now think of making the validity of that principle depend on the degree of accuracy of the experimental proof.[14]

Henri Poincaré (1854–1912), most profound of all, generalized this to all the laws of thermodynamics and to the Principle of Least Action as well:

> [these laws] represent the quintessence of innumerable observations. However, from their very generality results a consequence . . . namely, that they are no longer capable of verification.[15]

In other words, we must now simply take energy and its conservation as given. In fact, we would sooner invent new forms of energy than sacrifice the law of the conservation of energy. This is exactly what did happen in 1930. Wolfgang Pauli (1900–58) was troubled by problems with the β-decay spectrum and chose not to follow his predecessor Peter Debye's advice ('Oh, it's better not to think of it at all, like new taxes.'[16]) Instead, Pauli invented a new particle, the neutrino, as 'a desperate remedy to save . . . the law of conservation of energy'.[17] (Pauli describes his motivations in a strange letter that begins 'Dear Radioactive Ladies and Gentlemen'.)

Kinetic and Potential Energy—Why Just These Two?

Vis viva or kinetic energy, as it's now called, was first postulated by Leibniz in 1686, and then the idea of potential energy evolved gradually over the succeeding 100 years—an indication of the fact that potential energy is the subtler concept. It's unfortunate that the adjective 'potential' has stuck (this dates from Daniel Bernoulli in 1738), because it sounds as if the energy is not really there but is in abeyance. And indeed it is mysterious how a boulder can lie on a mountain top for a millennium with its energy seemingly locked away. But it is only the potential *kinetic* energy that is locked away (like snakes, we prioritize motion). The actual potential energy, the 'energy of configuration in space', is right there in a continuous 1,000-year-long agony of keeping the boulder away from the centre of the Earth. It's easy not to notice something that is always there—for example, the pressure on the soles of our feet or on our bottoms almost every hour of our lives. In fact, it took a Newton to notice and to put this into his law of 'Action and Reaction'.

But why are there just these two principal forms of energy? This is due to the fact that any given particle is in two states—it is isolated and interacting. In the latter state, the particle has an energy of configuration or, more generally, of interaction. This is the potential energy. Whether interacting or not, a particle still always has the energy proper to itself, and, if moving, this is the kinetic energy. The distinction between these forms of energy becomes blurred in modern field theories and, in any case, is not so antithetical as is sometimes assumed. Even in classical physics, the distribution of energy between these two forms depends on how the system has been modelled and on the choice of the frame of reference.

Imagine, for example, that we anchor our frame of reference to a wagon on the roller-coaster. We shall not be aware of our motion if we keep our eyes closed, but we do notice that we have to hold on tighter in certain sections and that our stomachs are wrenched (the interaction strength varies). Less colourfully, consider Kepler's description of planetary motion. The potential energy is proportional to mM/r as expected. But if in our choice of generalized coordinates we omit the angle, θ, from the kinematical variables, then there is an *extra* potential energy term that is proportional to $1/r^2$.

Another heuristic occurs in the case of moving charges.[18] In one frame of reference, a stationary electron is near a line of stationary electrons in a 'wire' (it's not really a wire, as we are ignoring the protons). There is a repulsive electrostatic force between the loose electron and the 'wire'. Viewed from a second frame of reference moving with constant speed v to the right (see Fig. 18.3), the electron and the 'wire' are now moving with speed v to the left. There is also a completely new effect (a new form of energy) that comes into play—there is an attractive magnetic force between the electron and the wire. But, and in beautiful accord with Einstein's Principle of Relativity, the same thing still happens in both frames of reference—the electron is repelled by the 'wire'. From the viewpoint of the second frame, the moving 'wire' has suffered a length contraction: the density of charge in the wire has increased, and to just such an extent that the extra electrostatic repulsion exactly counterbalances the magnetic attraction. In the second frame of reference we not only have the new effect, a magnetic potential energy, but the electron and the wire have kinetic energies that they didn't have before. (The explanation is heuristic because, strictly speaking, we should be talking of electric and magnetic field-strengths and not kinetic or potential energies. Note that the speed, v, isn't anywhere near the speed of light.)

There is still the question of which is more fundamental, kinetic or potential energy. Maxwell answers that it is the kinetic energy: it has only one form, $\frac{1}{2}mv^2$ (and a simple extension for rotations), whereas potential energy comes in a variety of forms, even while they all depend just on the relative configuration of the constituent parts. As Maxwell writes:

> we have acquired the notion of matter in motion, and know what is meant by the energy of that motion, [and] we are unable to conceive that any possible addition to our knowledge could explain the energy of motion or give us a more perfect knowledge of it than we have already.[19]

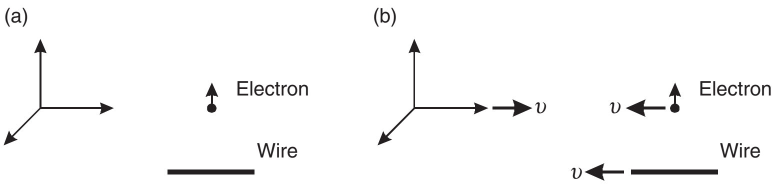

Fig. 18.3 An electron and a 'wire' seen from different frames of reference (redrawn from Duffin, *Electricity and Magnetism*).

On the other hand, with regard to potential energy, he writes:

> the progress of science is continually opening up new views of the forms and relations of different kinds of potential energy.[20]

This is true today: kinetic energy/energy proper to itself is of one type; potential/interaction energy is system-dependent and various.

Kinetic Energy—Where Does It Go?

Considering frames of reference again, it seems strange that the kinetic energy of an isolated body can disappear merely by switching from one frame of reference to another.

Take the billiard balls on the billiards table. Now pluck an individual ball out of this system and send it hurtling through space at high speed—say, 600 km s^{-1}. From a special frame of reference travelling with the ball, the ball's speed is zero. But we feel sure it still has energy—after all, it could land on someone's head and bring out a large bump.

The kinetic energy of this so-called isolated body seems to be crying out to be tied to *us*. Even our use of the phrase 'at rest' is loaded with anthropic connotations. And yet this elemental example is freakish in its simplicity: how often do we see an isolated billiard ball hurtling through space? We may paraphrase the legal maxim 'hard cases make bad law' into a physics maxim 'odd scenarios make bad intuition'. The temptation to bring ourselves into the system (of the *isolated* ball) is irresistible: we cannot truly grant that the billiard ball, all on its own, already comprises the entire system, and instead we demand to know where its kinetic energy has gone *relative to us*.

In fact, the billiard ball on the billiards table in front of me right now *is* moving at around 600 km s^{-1}, as that is the speed of our Milky Way galaxy through 'space'. I share this motion and therefore am completely uninterested in it: I am not aghast that this billiard ball has been 'robbed' of this vast store of kinetic energy but, rather, what is of concern to me is whether I can pot the red ball next go. If I can truly accept the system, the whole system and nothing but the system, with me either in it from the start or not there at all, then there is no paradox of energy suddenly going missing.

Rest Energy—an Absurdly Large Zero-point Energy?

It has always seemed puzzling to me that in changing frames of reference in the Galilean way (i.e. before Einstein), no account was taken of the mass. A frame was given a 'linear boost' (a spurt in speed), but it didn't make any difference whether that frame had masses in it or not. In Special Relativity this is rectified (by Einstein's Principle of Relativity and the requirement that the speed of light, c, is a constant for all observers). What comes out of it is that the zero-point of kinetic energy is not arbitrary any more—it is equal to the 'rest energy', $E_0 = m_0c^2$ (see Chapter 17; m_0 is the mass that an object has when stationary, i.e. in its 'rest frame'). Instead, the puzzle is now—why is the zero-point in energy so incredibly large? (It's an amazing 10^{17} times the typical everyday energy.) The answer is that the rest energy is somewhat like a store of energy. We are normally blissfully unaware of this vast store because, for one thing, we're usually dealing with scenarios nowhere near c and also the rest energy is tied up in a rather stable form as the rest mass of atomic nuclei. (This rest energy is a kind of potential energy: it depends on the configuration of particles within the nucleus and it trickles away—gets converted into heat—when radioactive nuclei decay.)

Magnetic energy is usually a very small fraction of the electrical energy and may often be considered as a relativistic correction to it. Likewise, the kinetic energy is usually a very small fraction of the rest mass energy and may also sometimes be considered as a relativistic correction.

The Rest Energy of Light?

We have been talking of how the energy of a material body is split between its rest energy and the energy of its motion (above and Chapter 17). In one special case, the case of pure radiation, the energy arises solely from the motion, as the rest mass (and hence the rest energy) is zero. The 'energy books' only balance if the speed of the motion has the highest possible value—the speed of light, of course.

What is Heat?

Heat is a new 'block' of energy and brings a completely new approach into physics: we can't answer the question of what it is just by reference to the macroscopic relations of classical thermodynamics, even those comprehensive statements such as the First and Second Laws. Heat is disordered energy. For a complete answer to what it is, we are compelled to delve into the (once speculative) microscopic arena. Then we find, as suspected all along by Bacon, Boyle, Hooke, Daniel Bernoulli, and others, that heat *is* the total effect of the random individual motions of an extremely large number of extremely small particles.

There are a number of qualifications to make. 'Random' doesn't mean lawless (the particles still obey Newtonian or quantum laws, as appropriate), but it means that the motions are not coordinated in a macroscopic sense. (So the *bulk* motion as occurs, say, in wind or in a sound wave, doesn't count as heat.) 'Particles' are almost always microscopic (for example, atoms, molecules, neutrons, photons, ions, electrons, fermions, Brownian particles, or combinations—as relevant to the investigation), but they could exceptionally be as large as stars.[21] The important thing is that the particles interact, somehow, else equilibrium could never be achieved. Finally, an 'extremely large number' means that there are so many of them that a statistical description is absolutely essential. A collection of, say, five molecules cannot manifest heat. The statistical outlook was a radically novel departure from what had gone before.

It appears as if merely scaling up the *number* can change the regime. This is incorrect—a true scaling would not change anything. The fact that we enter a new regime (in statistical mechanics) is due to the fact that we have not been able to guarantee a true scaling. We are now dealing with such a vast number of particles that it is impossible to give a complete description of the starting conditions—it would take more paper and ink than exist in the universe. So probability theory must be used. In classical mechanics—say, the collision between just two particles—probability theory is not necessary and the outcome is completely determined. ('Statistics' means having to do with a very large number of things; 'probability' means using the laws of chance.)

What of radiant heat? This can be modelled in two ways—as a collection of particles (photons) or as electromagnetic radiation. Either of these models leads to the same result: at equilibrium, the radiation has a characteristic distribution of wavelengths and the peak wavelength (the colour) is inversely proportional to the temperature. More typically, we could have a mixture of radiation and matter—say, light and electrons. When the radiation and the matter are in equilibrium with each other, then the hot electrons jostle around, have many collisions, and so continually change direction (accelerate) and thereby generate electromagnetic radiation. The equilibrium temperature of this generated radiation will be the *same* as the temperature of the hot electrons. The radiation in its turn will 'keep the electrons warm'—so the radiant and matter components of the system are maintaining the one consistent overall temperature. This is rather a simplified picture (a full treatment needs quantum electrodynamics and we have also ignored atomic transitions), but the gist is correct and we see that the concepts of equilibrium and temperature are creeping in. Before these are discussed, an old chestnut will be examined.

Does Heat Exist Only in Transit?

Classical thermodynamics is premised on the fact that a given state of a system can be completely specified by just a few macroscopic thermodynamic parameters. Whatever changes may be undergone, to whatever degree, and in whatever order (ignoring hysteresis effects), then, once the initial values are resumed, the system will have exactly the same colour, smell, sensation of warmth, viscosity, conductivity, taste, phase, or any other property. This goes for heat too—the system has exactly the same amount of heat in it (the internal heat) when the same state is resumed. On the other hand, the transferred heat, ΔQ, is variable and dependent on the route taken.

Now, here's the funny thing: thermodynamics makes a beeline for this variable, route-dependent quantity, ΔQ, while snubbing the fixed heat-in-a-body. The trouble is that the heat-in-a-body is not a very useful quantity for the very reason that it's too body-specific. Although we know that it is constant for a given state, we don't know what that constant value actually is: that would require microscopic theories and we are eschewing them in classical thermodynamics. ΔQ, on the other hand, while different for all the different routes, can at least be fairly easily tracked for all these routes. For example, the system under investigation could be surrounded by a water jacket and the total mass of water and its temperature change measured. Taking the specific heat of water as a given and assuming no heat leaks unaccounted for, then ΔQ can be determined.

While the internal heat is thus the less useful quantity, I wouldn't go so far as to say (as is sometimes claimed) that the heat-in-a-body doesn't exist: I don't know what's keeping my soup in the thermos flask warm if it isn't heat. This exaggerated claim is 'corroborated' by adding that there's also no such thing as work-in-a-body. Well, there *is* work-in-a-body—it's the total internal potential energy, for example, the sum of all the molecular attractions and repulsions. It is also a constant for a given state.

Ironically, in the one scenario so often exploited in thermodynamic demonstrations—the ideal gas—the internal heat *is* known. It is equal to the total internal energy and its value can be determined from the experimentally measurable quantity, C_V, the specific heat at constant volume.

Temperature

The fact that there are so many definitions of temperature, from widely differing arenas, and yet they all refer back to the same universal temperature, T, is *the* 'mystery' of thermodynamics.[22] Let us start by listing the multifarious definitions and occurrences of 'temperature':

- Sensation of hotness.
- The calorimetric equations, $\Delta Q = mc\Delta T$ at constant volume and $\Delta Q = m\lambda\Delta V$ at constant temperature (where c is the specific heat and λ is the expansion coefficient).
- Pippard's 'curves of sameness'.
- Celsius, Fahrenheit, and other empirical scales.
- The absolute temperature scale, T (in degrees Kelvin).
- The ideal gas law, $PV = nkT$.
- T in radiation laws, such as $\lambda \propto 1/T$ (Wien's Law) and energy flux-density $\propto T^4$ (the Stefan–Boltzmann Law).
- From the kinetic theory, $T \propto \langle \frac{1}{2}mv^2 \rangle$
- Boltzmann's factor, $e^{-(\beta\,\text{energy})}$, where $\beta = 1/kT$.
- $\frac{1}{2}kT = \langle\text{thermal energy}\rangle/(\text{degree of freedom})$.
- T in thermodynamic relations for reversible additions of heat: $dQ = TdS$; $\int dQ/T = \Delta S \geq 0$ and $dU = TdS - PdV$.
- T as an integrating factor for dQ.

How can we show that T is the same in all these cases? Well, we can't without writing a whole book just on temperature, but, in highly condensed form, we'll give it a go.

Granted that we have some primitive notion of a 'scale of hotness', we firm this up by calorimetry: adding 'heat' causes regular changes in 'hotness' and/or regular changes in some external parameter, typically volume. We define 'thermal contact' and 'equilibrium' and invoke the Zeroth Law. From all of the above, we can develop empirical scales of temperature (e.g. that ΔT is proportional to ΔV).

It must be admitted, however, that the calorimetry equations invite some circularity between ΔQ and ΔT. This always occurs when two new concepts are introduced at the same time (e.g. 'heat' and 'diathermal walls', or m and F in Newton's Second Law of Motion). There is no way out except that the innumerable experimental confirmations make the initial definitions consistent.

Continuing with the absolute temperature scale, we imagine using an ideal Carnot engine to map out the degrees. For every unit of work done, the temperature intervals are separated by exactly one absolute degree. Alternatively, the heat drawn out of or added to a reservoir during a Carnot cycle is proportional to the absolute temperature of that reservoir.

The immediate problem is that we don't have an ideal heat-engine at our disposal. However, we have the next best thing—a gas thermometer. This isn't a cheat: Carnot's proof shows that *any* engine type will do as long as we aim as closely as possible to the ideal, quasi-static conditions. The real gas thermometer is an excellent approximation to the ideal heat-engine and the approximation becomes better and better the closer the gas is to ideal (negligible intermolecular forces). So the inert gases (helium, neon, argon, etc.) are more ideal than other gases. In fact, the ideal case is reached for any gas in the limit as the density (pressure) of the gas is reduced to zero.

Two consequences follow (from this identification of real gas thermometers with ideal Carnot engines). First, the gas thermometer may be used to calibrate the absolute scale and to calibrate other types of thermometer that are more convenient in certain ranges of temperature (e.g. mercury-in-glass, platinum resistance thermometers, radiation pyrometers, adiabatic demagnetization, and so on). Secondly, the phenomenology of these gas thermometers—the gas laws of Boyle, Mariotte, Gay-Lussac, and successors—have ideal parameters. In other words, the T in '$PV = nkT$' coincides with the absolute T.

As regards radiation, provided that we consider a system in equilibrium, such as 'blackbody radiation' or radiation in a cavity, then we can talk meaningfully about its temperature. We can use electromagnetic theory to find the pressure of the radiation on the walls of the cavity and proceed with an analysis as if the cavity were a vessel containing a gas.[23] Or, equivalently, we can treat the radiation as an assembly of hot photons and apply the kinetic theory as if the cavity were a vessel containing gas molecules.[24]

Now, from the kinetic theory we find (and Daniel Bernoulli and his successors also found) that, at a fixed temperature, the momentum change due to individual collisions of molecules with the container walls leads to a macroscopic pressure and volume relationship: PV = constant (i.e. Boyle's or Mariotte's Law). In fact, in detail, the kinetic theory calculations show that $PV \propto \langle \tfrac{1}{2}mv^2 \rangle$, where P and V are macroscopic parameters and m and v relate to individual molecules. Comparing this to the ideal gas law, $PV \propto T$, we can then deduce that T is proportional to $\langle \frac{1}{2}mv^2 \rangle$, as Bernoulli and many others suspected.

But this is too quick. Huygens noted, in his analysis of elastic collisions and of the compound pendulum, that $\frac{1}{2}mv^2$ was coming out as a constant—a useful aid to calculation, but of no more significance than this (see Chapter 3). Now, 350 years later, we cannot be excused the same mistake as Huygens, and we must be alerted whenever any suspiciously constant thing like $\langle \frac{1}{2}mv^2 \rangle$ or 'average kinetic energy' comes into an equation.

We suspect that the deeper meaning of temperature and of its link with $\langle \frac{1}{2}mv^2 \rangle$ will have something to do with the idea of equilibrium. Also, we note that the directive 'heat flows from a high temperature to a lower temperature' is really subsumed in the principle that heat flows in the direction such as to minimize temperature differences (i.e. we don't get hot getting hotter and cold getting colder). In other words, heat flows in the direction leading towards temperature equalization; that is, leading towards equilibrium.

We shall explain the connection between temperature, kinetic energy, and equilibrium by following the line of thought given in Feynman's *Lectures on Physics*.[25] Consider a gas in a thermally insulating container and split into two compartments separated by an insulating wall that can slide without friction (Fig. 18.4). This dividing wall will get buffeted by the gas pressure from the left and from the right, and will eventually settle at a position such that the two opposing pressures are equal. Now, at this stage, equilibrium has been reached and the speed of the dividing wall is zero—or is it?

Equal pressures are consistent with an unequal ⟨kinetic energy per molecule⟩. For example, we could have a high density of slow-moving molecules on the left and a low density of fast-moving molecules on the right (considering just one type of molecule—say, nitrogen—on both sides). But this isn't yet the true, final equilibrium. Even after the pressure has been equalized, the molecules will continue to collide with the dividing wall *individually* and energy will be transferred such that slower molecules speed up and faster molecules slow down. Eventually, after many individual collisions, the true equilibrium will be reached and the molecules in both compartments will have a kinetic energy close to the average (e.g. around 6×10^{-21} J for a molecule of nitrogen at normal temperature and pressure). We say that the molecules' motion has been 'thermalized'. Moreover, the motion of the dividing wall itself has been thermalized (it is, after all, the agent via which thermalization

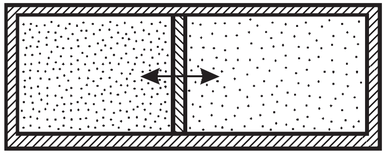

Fig. 18.4 Gas compartments separated by a moveable wall.

has been achieved between the two compartments). Its average kinetic energy will also be around 6×10^{-21} J and so, for a wall mass of say 0.2 kg, its average speed will be around 2.5×10^{-10} m s^{-1}—very small, but not identically zero. (The wall's average *velocity* is zero; in other words, its slow tremor is just as often to the right as to the left.)

Thus, at the microscopic level, the equilibrium is seen to be not static but dynamic (varying in time). The fact that time is involved is a clue: it reminds us that it has something to do with kinetic energy.

The phenomenon of pressure equalization is a necessary, but not a sufficient, condition for equilibrium. It happens sooner because it is a macroscopic phenomenon, like mixing colours with a broad brush. Only when 'colours' are mixed with a 'microscopic brush' will the true equilibrium be reached. The 'colour' is the kinetic energy per molecule (or per whatever thermal particle). Only once this molecular kinetic energy has been averaged out can we say that the *temperature* is the same in both compartments, and only then can we associate T with $\langle \frac{1}{2}mv^2 \rangle$. As we surmised earlier (see the end of Chapter 10), temperature is a more fundamental thermodynamic parameter than pressure.

Admittedly, there is some inconsistency in the thought experiment. Microscopic processes such as friction and heat conduction have been assumed to be negligible, but then the overall motion of the wall has been thermalized by nothing other than collisions with individual molecules. However, the thought experiment serves merely as a demonstration to explain a point. Its predictions can be checked up on by real experiments.

Let us nevertheless strive to make the thought experiment more and more realistic. For a start, no wall is truly insulating and the final equilibrium will, in practice, be achieved by heat conduction as well as by thermalization of the wall's bulk motion. Now let's introduce molecules of different masses—say, nitrogen in one compartment and helium in the other. Via the same process of billions upon billions of successive collisions, the averaging of the kinetic energy will be effected, irrespective of the different masses. (This mechanism has already been conceded in the thermalization of the massive wall itself.) After equilibrium has been reached, the lighter helium atoms will move correspondingly faster than the heavier N_2 molecules. However, the average kinetic energy per molecule will be the same, whatever the molecule's mass.

(At equilibrium, there will also be the same number of molecules per compartment volume, whatever the mass of these molecules, and this is the microscopic origin of Avogadro's Hypothesis.)

The kinetic energy we have considered so far is that due to the translational motion of the molecules (the energy that takes them from place to place within the volume of the gas) and is made up from three *independent* directions. Let us further consider individual molecules that can undergo various modes of internal vibration. We'll make an analogy that each nitrogen molecule is like a short spring connecting the two nitrogen atoms. The spring can squash up and stretch out, converting kinetic into potential energy and back again in a perpetual cycle—we don't need to worry about frictional losses, as we are already on the microscopic scale. On average (through time), each molecule will have half its vibrational energy as potential energy, half as kinetic—like a child on a frictionless swing (this is a standard result from the theory of harmonic oscillators).

Now, while vibrating, the molecule can crash into another molecule and transfer or receive kinetic energy (like bumper cars at the fairground). Eventually, after billions upon billions of such collisions, the vibrational kinetic energy per molecule will be thermalized (averaged out) with the translational kinetic energy in a given direction per molecule and with the vibrational potential energy per molecule. The same will be true even if some 'springs' have a different 'spring constant' from others (e.g. a stiffer spring could represent a more strongly bound diatomic molecule). A stiffer spring will not stretch out so far and will undergo faster oscillations—but its averaged out energy will be the same as for the looser spring. In other words, the spring constant serves somewhat the same role as the mass of a molecule—it affects the average speed but not the average thermal energy.

The analogy can be extended much further, as it turns out that *any* forces or potentials acting on or within a molecule can be mocked up by this analogical device of a spring with a certain spring constant. At equilibrium, all the different independent modes (degrees of freedom, or 'dofs') in which a molecule can absorb or release energy will then be thermalized within and between molecules. This is the famous 'equipartition of energy' of Boltzmann and others (see Chapter 17). Even a macroscopic external potential can be brought into this regime. For example, we could double the mass of the dividing wall or constrain its motion by an actual spring—either would be tantamount to introducing an external potential.*

The equipartition of thermal energy is at the same time both very difficult and completely obvious. It is difficult because it's not in the range of our experiences. In the same way as we never see one billiard ball all alone in space, we also never directly experience trillions upon trillions of events. It's hardly surprising, then, that much of heat studies, or of statistical physics, is counterintuitive.

On the other hand, we don't appear to find 'equilibrium' difficult to understand, and we readily accept the 'temperature' as a measure of it. It is instinctive that we wait a little time when taking someone's temperature, or stir the bath after adding hot water and so on. (When measuring the length of a table, on the other hand, we do it right away, with no hanging about.) Also, when told of the Zeroth Law of Thermodynamics (that if A is in equilibrium with B and with C, then B and C are in equilibrium with each other), we don't show amazement that no one has told us what A, B, and C actually are, or complain if A is mercury in an evacuated glass tube, B is a bath, and C is my armpit (well, I might complain).

The variable T is only going to work as a *universal* parameter if the thermal energy that it relates to is also universal. The equipartition theorem is asserting that there *is* such a basic, smallest unit of thermal energy and that, at equilibrium, it is spread between all the thermal dofs as evenly as possible. From the kinetic theory, Boltzmann found that this most basic unit was given by $\frac{1}{2}kT$ or a small multiple of this. k is a constant, known as Boltzmann's constant, and has the value $1.3806503 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}, so the basic unit is exceedingly small.

* We could also consider the gravitational potential, which will cause molecules in the atmosphere or the sea to have a density distribution that varies with height. Nevertheless, the average kinetic energy per molecule—that is, the temperature—is the *same* regardless of height. (This is counter to experience, but only because we never see the sea or the atmosphere under equilibrium conditions: the Sun provides heat from above in the day, the Earth provides heat from below, and beyond the atmosphere it's very cold.)

Although we have given an example relating to a gas, the astounding fact is that this basic thermal energy unit is the same whatever the system (e.g. a wire, a surface film, an electrical cell, or a paramagnetic solid[26]). In the case of radiation, as we have stated earlier, thermalization occurs between the radiation and any matter particles that are present; and at equilibrium the matter and radiation components have the same temperature. (For pure radiation in empty space, there are an infinite number of dofs (the electromagnetic waves can take up any wavelength), and so the radiant energy is spread between many dofs and the rise of temperature with energy is correspondingly slow (for radiation, $\Delta Q \propto T^4$; whereas for matter, $\Delta Q \propto T$). However, the basic principle is still the same—the heat energy comes in the smallest units dependent on the temperature.)

The reason *why* there is only one fundamental thermal unit must relate ultimately to the fact that all systems are microscopic *in the same way* and thermalization occurs by microscopic collisions *in the same way*.*

In other words, heat is like a house made up from unit-sized 'bricks'. The bricks are thermal energy rather than matter and are the same whatever the design of house (hydrostatic fluid, paramagnetic solid, electric cell, surface film, radiation, etc.). However, unlike the matter-house, in a house-of-heat the size of the elemental brick actually depends on the temperature—the higher the temperature, the bigger the brick (the brick size equals $\frac{1}{2}kT$).

Specify the temperature and you have specified what the elemental heat unit is. You have then said *everything that needs to be said* (e.g. how two water-baths come to a common temperature, what happens when heat and work is added to a system reversibly, etc.). There is only one β in the Boltzmann factor and T acts as a 'universal integrating parameter' for dQ. This amazing versatility is possible because the parameter T already implies equilibrium conditions, and because the microscopic thermalization process is universal.

The flaws in an analogy are just as enlightening as the points of agreement. Not only does the elemental brick depend on the temperature, but the brick size is only an average—and the higher the temperature, the greater is the *spread* in brick size away from this average.

* We are avoiding special regimes such as ultra-low temperatures or very high densities.

This is all very well, but is any of it true? Let's look again at the Maxwell–Boltzmann distribution of molecule energies in a gas at equilibrium (see Chapter 17). Remember that the equilibrium is *dynamic*, so energy is constantly being transferred from one molecule to another. However, the probability of each and any smallest increment in energy is the same. A large deviation from the average energy is like a succession of smallest increments. For independent events, the probabilities always multiply (this is a standard result in probability theory). Thus for larger and larger deviations, more and more small probabilities are multiplied together—therefore the distribution will tail off exponentially. So, the Maxwell–Boltzmann distribution is *exactly* what is expected, given the assumption of equal probabilities or randomization.

What is more, the Maxwell–Boltzmann distribution has been experimentally checked (e.g. by using a rotating slit that allows molecules only within a narrow range of speeds to reach a detector). The Maxwell–Boltzmann curves for a few temperatures are shown in Fig. 18.5.

To summarize, the reasons why there is just one universal temperature are:

Universality of T

- *T*, by its very definition, only makes sense at equilibrium.
- Thermalization occurs in one fundamental way—by random, microscopic collisions.
- The molecules (or other entities) go around singly and not in gangs (of, say, millions or billions).

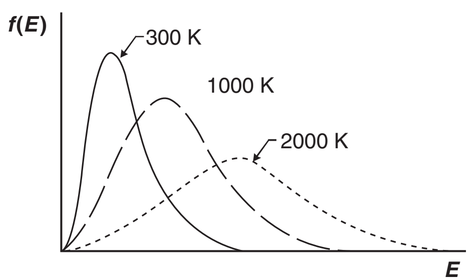

Fig. 18.5 Maxwell–Boltzmann curves (not normalized to each other).

- There is only *one* equilibrium parameter, the temperature—we are not trying also to equilibrate 'femaleness' or some other attribute.

Thus, for the concepts of heat and temperature, we have found agreement between the microscopic theory, the macroscopic theory (classical thermodynamics), and experiment.

Why, then, are not heat and temperature sufficient to explain everything? Why do we have need of yet another parameter, the entropy?

Entropy

The short answer is that many physical processes continue even when T is constant—for example, melting, boiling, and the free expansion of a gas. Before continuing with this, we shall honour our promise of Chapter 17 to try to explain Boltzmann's microscopic definition of entropy as $S = k \log W$. We cannot prove that this combinatorial approach is justified but we can show that it's not as far-fetched as it might appear.

First, we remind readers that this sort of thing has happened before. When a coin is tossed, there are two entirely different ways in which we can find out whether it will come up tails. We can either solve the dynamical problem (try to track the trajectory of the coin by solving the equations of motion) or we can use probability theory. Now in a typical, everyday scenario where entropy is involved, there are 10^{24} or more dynamical variables (e.g. this is the number of molecules in 22 L of air). To solve the dynamical problem is out of the question. Even to write down the starting conditions would take more time, paper, ink, or computer memory than exists in the universe.

Secondly, the Maxwell–Boltzmann distribution was itself derived by a probability-theory-down rather than a collision-theory-up analysis (Chapter 17).

Thirdly, the thermodynamic parameters of state, P, V, U, and T, all relate to averaged out features. Boltzmann realized that if irreversibility was ever going to be explained, then the parameter S was the one to do it ($S > 0$ is the only macroscopic formula relating to irreversible processes) and S would therefore have to 'bite the bullet' and describe microscopic attributes.

In his microscopic theory, Boltzmann posited that there are a large number of finely spaced energy levels and that these can be populated by the various microscopic (energetic) degrees of freedom (dofs). A register is made of all the ways that this can be done. It is assumed that there is an equal probability that any dof can occupy any energy level (consistent with the requirement that the total energy is a constant). This is a basic tenet of statistical mechanics, like assuming that a die can land with equal probability on any of its six faces. Without some such assumption, probability theory can't even begin. It follows that the probability of a given macroscopic distribution of, say, 10^{24} dofs is just the number of microscopic ways in which this macroscopic state can be achieved.

Now as the dofs (say, 'Tom', 'Dick', 'Harry', etc.) are indistinguishable, then permuting the 10^{24} different names makes no difference to the overall energy profile. However, each permutation still counts as a distinct microstate ('complexion'[27]) of the system with that energy profile. Consider two absurd and extreme energy profiles (quantum mechanics has been ignored): (a) one dof has all the heat energy—the remaining dofs have zero energy; (b) all dofs have the same energy. For 10^{24} dofs then (a) can be arrived at in 10^{24} ways and profile (b) can be arrived at in $(10^{24})!$ ways. So the profile (b) will occur $(10^{23})!$ times as often as profile (a).* Not only is the second distribution overwhelmingly more probable ($(10^{23})!$ is an absurdly large number, bigger than a googolplex) it is also completely bland and featureless (10^{24} equal chunks of energy), whereas the first case has just one crazy dof careering about.

Thus, just from combinatorics, the most uniform distribution is overwhelmingly the most probable. But this is exactly what is required: S keeps on increasing, and at the maximum the distribution is the most probable (greatest statistical weight) and the energy has been spread as evenly as possible—it has been dissipated.

Note that Boltzmann's combinatorics and the Maxwell–Boltzmann distribution both hinge on exactly the same crucial assumption of randomness or equal probability. Each approach therefore supports the other.

The word 'probability' is misleading: it suggests that S is a pure number and has a value somewhere between 0 and 1. However, the statistical weight is not really a probability: it is actually a head count (or, I should say, an energy count) and S has the appropriate dimensions of energy for a given temperature (because of being multiplied by k). As energy is an extensive quantity, so entropy is an extensive quantity.

* $x!$ is read 'x factorial' and is equal to $x(x-1)(x-2)\ldots 3.2.1$. Never mind if this is new to you—just know that $(10^{23})!$ is an incredibly large number.

(The logarithm function is used for reasons of mathematical ease. It converts a product of exponents into a sum and it is a much more slowly varying function than an exponential.)

Now, when heat is added to a system there are two main ways in which the statistical weight and hence the entropy can change. (1) The energy levels stay the same but their occupancy changes. This is purely thermal: heat is added, T increases, and so some higher energy levels may be accessible for the first time. (2) The energy levels themselves—their actual architecture—changes. This happens when, say, the volume changes, that is, work has been done. From both causes, pure heating or work, the statistical weight and hence S will change—in fact, they'll increase. All this ties in with the phenomenology and with classical thermodynamics.

The formula $S = k \log W$ looks disarmingly simple—how can it have so much physics in it, applicable to all different set-ups, such as chemical mixtures, soap films, electrolytic cells, and so on? But, just as with the Lagrangian (Chapter 7) and the Hamiltonian (Chapter 13), all the physics is implicit in the specifications—what the dynamical variables are, what the energy levels are, the boundary conditions, and so on. So while W gives the statistical weight for a given system, the physicist must first of all do all the hard thinking and specify the actual energy-level architecture for this system. To give an analogy, S might be the layout of streets, the 'A to Z' of London, and then the 'temperature' could be the average speed of the cars, equilibrated by driver-interactions.

(One final question concerns the absolute value of S—is it possible to determine this, or is it only possible to determine differences in entropy? The answer is that before quantum mechanics, entropy could only be fixed to within an arbitrary constant. After quantum mechanics, the statistical weight for certain states could be determined absolutely. It was found that the absolute zero in temperature could never be reached and that as zero was approached the entropy approached zero as well. This is Nernst's Law, also known as the Third Law of Thermodynamics.)

Let's assume that you are now happy enough to accept Boltzmann's microscopic formulation of S. We return to our starting question—why entropy?

Gibbs started his great work of 1878 with the words:

> . . . when the entropy of the system has reached a maximum, the system will be in a state of equilibrium. Although this principle has by no means escaped the attention of physicists, its importance does not appear to have been duly appreciated.[28]

Gibbs was drawing attention to the fact that when a function is at an extremum value (e.g. a maximum), then it is stationary with respect to small changes. Thus, when the entropy has been maximized and we are at equilibrium, the system is stationary with respect to infinitesimal changes in the distribution of energy. This, finally, is the driving force behind all statistical processes in physics—the system is striving to reach and maintain (dynamic) equilibrium, which is the same thing as saying that the energy is trying to get as *evenly distributed* as possible.

We can at last explain why heat must flow from hot to cold. As the basic unit of thermal energy is proportional to kT, then this unit is smaller as the temperature is lower. The energy can therefore be distributed more finely at lower temperatures.

But even when the temperature has reached its equilibrium value, then another process takes over the task of smoothing out the energy. Take the mysterious example of the free expansion of a gas (see Chapter 10, Part II). The volume increases and so the microscopic energy levels become more numerous and closer together. (Quantum mechanics is needed to explain this, but it's a bit like standing waves on a rope: as we consider longer and longer ropes, then there are more and more possible standing wavelengths, and so the possible energies are more and more crowded together.) As there are more and more finely graded energy levels, then the statistical weight and hence the entropy has increased and the energy is distributed more evenly.

To sum up, temperature equilibration smoothes out the energy as best it can, but when a common temperature has been reached then entropy maximization carries on with the job and leads to an energy distribution that is smooth to *greater and greater precision*.

The Second Law of Thermodynamics and Irreversibility

In 1909, the Greek mathematician Constantin Carathéodory (1873–1950) finally succeeded in ridding the Second Law of constructs such as heat-engines and putting it into its most economical form:

> In the neighbourhood of any equilibrium state of a system there are states which are inaccessible by an adiabatic process.[29]

'Neighbourhood' is reminding us that entropy increases locally and incrementally—we can't save up the increase for later or for somewhere else. 'Adiabatic' means 'without transfer of heat' (see Chapter 10), so there always exist close states that are *only* accessible *with* heating. In other words, work isn't enough to undo or set up the sort of messing that only 'heat' (randomization) can do. For example, when doing the housework, I can never clean away *all* the dust but also, by work alone, I can't even set up the initial arrangement of dust—one grain here, another grain there, and so on.

Maxwell was the first person to introduce the use of probability theory into physics and was also the first to appreciate that the irreversible nature of the Second Law of Thermodynamics was statistical in origin. (We have already quoted Maxwell's poetic example of raindrops flying upwards and people becoming unborn; see Chapter 17.)

The fact that we *don't* see time running backwards is therefore not because this would contravene any laws of physics, but only because the starting conditions in these time-reversed cases are very specific—they are incredibly unlikely to occur at random. The increase in entropy is not towards or away from a particular state, but only towards increased randomness. It's not the fault of entropy that out of a googolplex's worth of starting states and a googolplex's worth of final states, only a handful are special—to us.

For example, think of the number of ways in which colours can be spread between pixels in a photograph. For a random distribution, then billions of different photos will come out uniformly brown. The photo with my daughter and her cat will be lost amongst this crowd. I can't help attaching special significance to this photo, but the entropy 'nose' is only tuned to various extensive aspects—in this example, to the distribution and conservation of the total amount of ink.

Loschmidt, in the 1860s, considered the same time-reversed starting conditions as Thomson, but thought that the outcome was paradoxical—if time had run backwards then entropy would have *decreased* and the Second Law would have been contradicted. Boltzmann and Loschmidt debated this in friendly fashion for years and Boltzmann came round to understanding what Maxwell had known all along: the Second Law isn't true in an absolute sense—decreases in S are not ruled out, but they become vanishingly unlikely as the number of particles involved increases. The very large number of microscopic components in any typical macroscopic setting (i.e. in real life) means that departures from the Second Law never happen *in practice*.

You may be wondering how 'design' and complexity can ever be achieved in a world heading inexorably towards randomness. What happens is that 'designer features' are arrived at by a process of statistical default—like the grading of gravel into different layers as a barrel is bounced around on a truck, or the drawing that emerges as sand dribbles out of a hole in a swinging bucket. Sometimes bits of design are temporarily frozen in (stars, planets, life, and so on). A more humdrum example is the child's toy where lots of little ball bearings can be jiggled into lots of little hollows.

The process is reminiscent of Darwin's theory of evolution, where 'design' is arrived at by a long succession of very small random mutations. Those who baulk at this would despair at the even more stringent constraint in all of physics: the driving force behind all physical change is increased randomization, and every local increase in order must be paid for by an even larger increase in overall disorder.

Some processes seem to defy the Second Law and must be examined carefully. In photosynthesis, gaseous molecules of CO_2 and H_2O are tamed and made to reside quietly together in large structures (such as cellulose and carbohydrates), with only the release of gaseous oxygen to compensate. Worse still, sunlight is absorbed. Ingo Müller[30] suggests that the plant must simultaneously release 500 times as much water vapour for leaf-cooling as it absorbs for leaf-growing. The random motion of these water vapour molecules compensates for the increased order implied in the highly structured, large organic molecules.

The one 'paradox' left to be explained is how macroscopic irreversibility can come out of microscopically reversible laws. There is an overall direction to time (people grow old, the strains of Beethoven's Ninth Symphony turn to heat, etc.), but the laws governing individual collisions (Newtonian mechanics) are not affected when t is replaced by $-t$.

We have already given the answer to this paradox: it is that the asymmetry between the initial and final times occurs not because of any asymmetry in the laws of physics, but because of a pre-existing asymmetry in the initial and final states. In other words, the initial state (of the universe) is rather special—it is far from equilibrium. The question then becomes: why is the initial state far from equilibrium? There are two possibilities: either the current state of the universe represents a large statistical fluctuation or the universe started in a special way. Physicists have adopted the second option ever since the 'Big Bang' theory.

19

Conclusions

To return to the physicist at the start of this book, she can now explain energy as the 'go' of the universe, as what makes things happen. She can further explain that there are two main forms of energy—kinetic, the energy of movement, and potential, the energy of interaction—and that there is a continual, flowing interchange between these forms, a perpetual 'dance to the music of time', as exemplified in Nicolas Poussin's magnificent painting (frontispiece). But the kinetic component is the more fundamental—it is of one basic type and is also the mechanism by which statistical energy reaches equilibrium. Finally, there are three laws governing energy: it is conserved within a closed system; its distribution is randomized as much as possible, and as finely as possible; and, with time, as action, it is minimized.

The physicist must stress that energy is not a 'vital force' or 'life force'. When talking with my non-scientific friends, I am interested to find that this misconception is still pervasive, even 160 years after Helmholtz's quest to banish it.

Feynman's analogy explains that in order to know what energy is, we must discover the specific and various mathematical formulae that will preserve the constancy of the very thing we are looking for. But there is another angle: we could instead ask—what makes things happen? Physics has two quite different approaches to this question, the energy view and the force view. We will compare these two in a while, after reviewing the historical progression.

'Energy' was not simply waiting to be discovered, in the way that a palaeontologist might find the first ichthyosaur, or a prospector might stumble across the Koh-i-noor diamond (these discoveries were contemporaneous with the discovery of energy in the middle of the nineteenth century). The concept of energy had to be forged, and out of ingredients both physical and metaphysical.

First, we saw how the futile quest for perpetually acting machines hinted that *something* was being conserved. Then Leibniz identified 'mv^2' (*vis viva*, or 'living force') as that something—except that it was only conserved in those collisions where it was conserved. There followed a 50-year controversy over which was the important measure of 'motion'—Newton's momentum or Leibniz's *vis viva*.

Amontons built a small-scale prototype 'fire-mill', defined 'work' for the first time, and was able to quantify and compare the work capabilities of men, horses, and his hypothetical heat-engine.

Then, the concept of potential energy began its gradual evolution and also its merging with the engineer's 'work'. One family, the Bernoullis from Switzerland, discovered that kinetic and potential energy were conserved jointly—as what we now call mechanical energy.

In the last quarter of the eighteenth century, the mechanics of Lagrange, the acme of analytical perfection, brought in a new outlook—not just conservation, but also the most economical path in time.

Then, a most surprising thing happened. On the one hand, Lagrange's mechanics was successfully applied to everything (everything *mechanical*) and especially celestial motions; on the other hand, the Scottish engineer, Watt, invented a most improbable and revolutionary engine—the steam engine—and in less than half a century steam engines were pulling locomotives and powering the Industrial Revolution; and, on the other other hand again, the French chemist, Lavoisier, developed a theory of heat. But (and this is the real surprise) these three domains—the conservation of energy in 'clockwork' mechanisms, the power revolution, and the theory of heat—all remained completely separate from one another. So, while the countryside began to be criss-crossed by steam trains on railway lines, 'energy' still had not been discovered.

One crucial advance was the progression from a statical to a dynamical viewpoint: from a static balance of forces to the constant interchange between different forms of energy; from the static heat-fluid finding its level to the dynamic equilibrium of Prévost; from the static to the dynamic descriptions of a gas. As Thomson wrote in 1872, 'I learned from Joule the dynamical theory of heat and was forced to abandon at once many, & gradually from year to year, all other statical preconceptions.'[1]

The caloric theory of heat (that heat was a weightless and invisible substance) had its roots in chemistry. Chemical substances were invariably conserved, so heat should be conserved too. The theory was remarkably successful at explaining almost everything—from the behaviour of gases to the three phases of matter and the speed of sound.

It is hard to 'un-know' what we know now (that a gas is made up of myriad molecules in a perpetual whir of motion) and so it is mystifying how gases could ever have been explained on the old theory. It is especially dumbfounding that gas pressures could arise from this totally static view of a gas (see Figs 10.1 and 10.2).

There were really only a few niggling details that the caloric theory couldn't explain convincingly: the 'inexhaustible' amount of heat that could be generated by friction, the transmission of smells across a room, and the sublimation of ice into a vacuum.

It must not be thought, however, that the impetus for the final discovery of energy came from any such discrepancies. Like the Copernican theory and Einstein's theory of General Relativity, the new theories were not motivated by anomalous observations but by their superior explanatory power, their beauty and 'simplicity', and the range of their reach. It was in this spirit that Mayer and Helmholtz appealed to the age-old wisdom that 'cause equals effect', and Joule to the 'Creator', and all finally understood that a collision couldn't simply result in the cessation of motion of the colliding bodies without the *energy* of that motion having gone somewhere; an animal couldn't do mechanical work from some perpetual fount or vital force; a gas expanding against a piston and raising weights was doing work, expending *energy*, and this had to be sourced from the heat *energy* of the gas.

These were hard steps to take, as they required not only a break with the ruling caloric theory but the overriding of a category error: heat and work were as different as p.s.i. (pounds per square inch) and psa (pleasant Sunday afternoon).

Nevertheless, niggling details don't go away and, moreover, theories are ultimately differentiated by these details—how else could we tell which one was right? So it was essential for all the seemingly insignificant minutiae, whether in everyday observations or in the course of an experiment, to be noticed and explained: the drift of a smell, the direction of comets' tails; the rate of melting of snow; the air temperature during a thaw; shadows being unaffected by breezes; the precession of planetary orbits; a warming effect beyond the red end of the spectrum; drops of colourless liquid in a flask; the rate at which a magnet is moved towards a coil; the dip and orientation of a compass needle; sublimation in a vacuum; bubbles at electrodes and sparks at commutators; where cannon-balls land relative to the mast of a ship or the Tower of Pisa; the heating of carriage wheels and axles; the impossibility of reflecting 'cold'; the viscosity of gases; the clicking sound as billiard balls collide; the redness of blood in the Tropics; . . .

Daniel Bernoulli was, I think, the first to understand the idea of energy-in-the-round. In 1738, in his *Hydrodynamica*, and in his later works, he talked not only of the conservation of '*vis viva*' in a variety of mechanical settings, terrestrial and planetary, but also of the *vis viva* in a swirling stream, an expanding gas, a lump of coal, steam, a curved metallic band, and a cubic foot of gunpowder.

By the 1860s there was no doubt about it, and Maxwell was able to observe that, in the conduction of heat, the molecules stayed in their same average locations but *energy* was transmitted through the gas.

Why did the material theory of heat not cause more divergence with experiments (apart from niggling things)? There was the ever-adaptable fudge-factor, the latent heat, but, more generally, there was the fact that, for the most part, thermodynamics dealt with *equilibrium* conditions and with the *averaging* of quantities, such as pressure and temperature. But these quantities didn't then pick out the details that could generate a conflict. The temperature, in particular, being universal, could be embedded in almost any theory, or, to put it another way, temperature is the parameter that shows that *energy's* the thing.

It is pertinent to ask, then, how any atomic details (e.g. atomic mass and size) could ever be extracted? The answer is that these details have invariably been determined from measurements off the average and/or off equilibrium (through processes such as diffusion, conduction of heat, 'random walks', etc.).

Not so much a niggling detail as a blindingly obvious one—with hindsight—was the observation that heat always flows from hot to cold. Also: rivers flow downhill, projectiles and pendulums slow down, engine efficiencies are always less than 100%, and time goes from the past to the future. This universal evidence would culminate in a universal law—that energy is distributed in the most random (disordered) way possible. This was the first time that an absolute direction came into physics.

There were thus two over-arching laws of thermodynamics: the First Law—energy is conserved—and the Second Law, entropy always increases. These laws implied that machines with greater than 100% efficiency were impossible (the First Law) and even machines of exactly 100% efficiency were impossible in practice (the Second Law). This has been parodied as 'you can never win' and 'you can't even break even'.

In spite of Feynman's allegory, it turns out that energy is too important to leave to just those systems where it is conserved. Hamilton extended Lagrange's analysis to include such non-conservative cases. This wasn't to say that, all of a sudden, the law of the conservation of energy didn't apply but, rather, that Hamilton was able to treat systems in which the potential energy did depend on time in a systematic and predictable way—say, the rug is slowly being pulled out from under my feet.

The advance of Hamilton's approach over Newton's was that it merged mechanics (dealing just with particles) and optics. It was therefore especially amenable for adapting to the wave-particle duality of quantum mechanics. Hamilton's spread of possibilities in phase space became the distribution of *probabilities* for the position and momentum of a quantum-mechanical particle; and the function called the 'Hamiltonian' in the new physics was the analogue of the 'work function' (the sum of kinetic and potential energy terms) in the old physics.

Also, Hamilton's succession of infinitesimal transformations meant that disturbances and interactions could be propagated *locally*; in other words, by inspection of just the local neighbourhood. Hamilton's mechanics could therefore be readily extended to the cases of continuous media, statistical and quantum mechanics, and gravitation, as in all these cases a local or 'field' description of nature is appropriate. The field may be likened to an eponymous field of grass in which every blade has a specified position and various attributes (say, the height and direction of the blade-tip). Without these attributes the field isn't defined (just as a field no longer exists in the grass piled up in a haystack). More than just a question of mathematical expediency, it turned out that the field description was essential, was borne out by experiment, and that the field contained energy.

Hamilton's mechanics was also a step towards Einstein's Theory of General Relativity. Two hundred years earlier, Johann Bernoulli had the pregnant idea of using an optical analogy for a particle falling under gravity. The particle took the least time and its path was like that of light going through a medium with decreasing refractive index. Now, in Einstein's theory, not only light but all particles in free fall near a large gravitating object take the 'least path'. Mass (strictly, mass-energy) is not only a cause of curvature but can be an outgrowth of it. This last is an astounding result but, as we have seen, there were already intimations of it in d'Alembert's Principle.

Which is better, Newton's 'force view' or Hamilton's and then Einstein's 'energy view'?

'Force' is a more intuitively comprehensible concept than 'energy'—it has direction and it acts on an individual body. On the other hand, potential energy relates to the configuration of a collection of bodies—a system. But, if the system comprises just one body, what does configuration mean and what are the systems aspects of the kinetic energy of this one body? All is resolved when we come to realize that the body moves in space and time and, in the energy view, these are not blank, absolute arenas for individualistic behaviour but, rather, every 'position' (including the time coordinate) is an 'event' and these events are *part of the system*. The sharp distinction between force, matter, space, and time has become increasingly blurred with each step away from Newton and on to d'Alembert, Lagrange, Hamilton, Maxwell, and Einstein. In other words, the force picture seeks to explain, separately, what there is and what happens, whereas in the energy view, all is explained in one go.

Thus there may be 'configurational aspects' for even just one body in space and time; and in continuous media (think of blancmange), space is seeded with mass-points, each in a state of stress. In electromagnetism, the mass-points are moving and charged and, as Faraday found, the field spreads from the body into the surrounding space. And even 'empty space' may be groaning with stresses and strains, as in the stress–energy tensor of General Relativity.

The exemplar of Newtonian mechanics—one body deflected or accelerated by a force—is such a simple scenario that the systems aspects are not readily apparent and are a seemingly redundant complication. The force-picture continues to be of enormous utility but is not sufficient to cover all cases, and we now need the complexity of the systems view. In fact, we saw how d'Alembert's Principle tried to bend Newton's forces to the systems approach even before energy had been discovered: the reactive forces *act in concert* to counterbalance the externally applied forces.

Thus it is the systems view, the *energy* view, which is more generally applicable and more fundamental. Within the system, each block of energy maintains its specific mathematical formulation but there is the possibility of the transformation of one block into another, a continual interplay between the blocks, a perpetual 'dance to the music of time'.[2] However, there is a price to pay for the versatility and universal applicability of the systems approach—'energy' is more varied, complicated, and less intuitively comprehensible than 'force'.

The energy approach is not just more versatile, it is more profound, as the problems of absolute space, time, force, and inertial mass are all avoided—everything is now relative to the system. No longer are there hiccoughs when, say, the kinetic energy disappears between frames of reference. Also, in Einstein's Theory of General Relativity, 'uniform motion' has been extended to include acceleration, and instead of hypothesizing an absolute space and time we can take the presence of a large gravitating mass as a given (and require only that space and time are 'flat' when examined sufficiently locally). Einstein's energy/systems view is more complex than Newton's force of gravity, but it is more 'respectable' (needs less hypothesis)—after all, all our observations really have been made near a large centre of gravitation.

In fact, the 'energy principle' and the Principle of Relativity are the two most important principles in the whole of physics. But, in the second case, how can such a metaphysical principle have anything physical to say? Answer: its physicality lies in the fact that it enables the true universal aspects to be distinguished from the mere artifice of a particular point of view. The two principles acting together can be sloganized as follows: 'the system is all, thereafter everything is relative'.

So, as the relative kinetic and potential energy contributions may change in going from one observer to another, what is that kernel of physical reality that can be plucked out of these shifting sands? It is not the mass, nor the length of a ruler, nor the tick of a clock, nor even necessarily the momentum or energy. It is that more complicated thing, the 'minimum of the action taken through time'.

'Action' is defined as 'energy × time', so we see that energy is still at the heart of physics, but 'action' is seen to be an even more comprehensive concept, as it spans the whole dynamical 'space' of the given problem. Why, then, do we have more to do with energy than with action in physics and in everyday life? It is perhaps because action is too high-level a measure to be a useful physical handle for most purposes—like saying 'the answer is 42'. We are nevertheless bound to use this high-level parameter, especially in quantum mechanics, when we don't have complete information or the system is too complicated for a more detailed analysis.

Energy has intensive (temperature) and extensive (entropy) attributes. While action may be divided up in two ways, as the 'conjugate variables' (p,q) or (E,t), it is in the latter pairing where the extensive and intensive aspects are most polarized. But there appears to be some overlap in the duty of the parameters 'energy' and 'time'.

For example, consider an isolated body satisfying Newton's First Law of Motion: it is travelling alone in 'outer space' at constant speed and in a straight line. But what if, all of a sudden, it spontaneously combusts, like a haystack, or it decays, like a radioactive nucleus? We shall then have to admit that either the original system wasn't truly isolated (an external force came along) or that the body had internal structure (internal potential energy) of which we were previously unaware. Next time round, we shall have to keep on and on watching the body to make sure that it really is isolated. Only when time is uniform and featureless can we be sure that energy is conserved; but only if energy is conserved can we be sure that time is the same everywhen, and not marked or bunched up or stretched out.

Consider another impossible thought experiment. We are on a ship and watching an infinitely long wave go by our porthole (the wave could represent a free, quantum-mechanical particle whose position is completely undetermined). If we close the curtain to a mere crack, then we shall know exactly when a crest of the wave goes by, but we shall be totally ignorant of the frequency of the wave (related to the particle's energy). If, on the other hand, we want to be completely certain of the frequency, then we shall have to watch for an indefinitely long window of time (note, this explanation is heuristic).

It has been a recurring theme that the experimentation, the mathematics, the technology, and the ideas themselves all had to push forward together in order for the concept of energy to emerge. The biggest philosophical shift was the realization that nature must be inspected and *measured*: it was not sufficient just to learn from authority (chiefly, the authority of Aristotle). Numbers had to be mapped on to physical quantities, as occurred in any measurement and in the setting up of the Cartesian coordinate system. 'Time' was at last understood as something that could be brought into physical law, and compared with 'distance'—but how could motion ever begin? This couldn't be resolved until Newton and Leibniz had invented the calculus.

We have seen how important it was to quantify, to carry out precision experiments, and to take notice of any details and 'whimsical' observations. In Descartes' and d'Alembert's rational approach, the science was empirically based only to the extent that a nod was given to 'well-known' experimental facts, such as that a heavy body was deflected less than a light body after a collision, and so on. This would never do.

The steam age both caused and followed from the Industrial Revolution, requiring advances in mining, metallurgy, boilers, machining, and general engineering. The study of thermodynamics required precision thermometry, which in itself entailed strong glassware, tubes of accurate bore, pumps, seals, valves, the collection and storage of gases, and so on. The reader can continue adding to this list when it comes to the voltaic pile, electrochemistry, electromagnets, optics, and so on.

New technology answered old questions, but always brought in new questions. For example, the telescope and microscope provided more stringent tests of theories, but also opened up new vistas that required philosophical readjustments, such as when Galileo's telescope revealed craters on the Moon (the sub-lunary sphere was thus imperfect) and the microscope revealed Brownian motion. In this last case, Thomson's paradox of a pendulum coming to rest in water would be shown to need revision: the pendulum never comes to a complete standstill, but executes miniscule thermal motions.

As the physical thing to be described became more and more complex, so the mathematical 'object' needed for its description became more and more complex. Instantaneous velocity was a new object, $d\mathbf{r}/dt$, and not the same as $\Delta r/\Delta t$. Some other mathematical objects that were required were $\frac{1}{2}mv^2$, $\int\mathbf{F}.d\mathbf{r}$, the Laplacian, the 'variation of $\int(T - V)\,dt$', the stress–energy tensor, and so on.

The mathematics didn't just apply to things but to processes as well; for example, to the instantaneous changes in direction when hard bodies collide, to the quasi-static processes, to the statistical treatment of a large number of things, and so on. Nature had to be hypothetically sliced, chopped, stretched, smoothed, and generally beaten into mathematical shape. It was Galileo who first understood the idea of an idealized experiment and how to winkle out an ideal law from real observations. One of the biggest advances (and consequent philosophical adjustments) concerned the admission of probability into physics—in the treatment of 'errors', equilibrium processes, and quantum mechanics.

All told, the lesson from history is that if the physics is right, then the philosophy is right—even if it takes the philosophers one or two hundred years to come round.

The history also shows us that we forget our victories as soon as we have won them. But we are all physicists now—yes, even those of us who would least like to think so. Consider everyone's intuitive understanding of gravity—something that makes apples fall to Earth and planets orbit around the Sun: this intuition is actually a remarkable and sophisticated intellectual understanding. As Hamilton said, 'Do you think that we *see* the attraction of the planets? We scarcely see their orbits.'[3]

In other words, physics has permanently changed the view, changed what it means to be human. It is true that our physics is not absolute but must continually be corrected. However, it has passed into the realm of *knowledge* and is not nearly as provisional as is commonly made out.[4] We have seen how it has been discovered in so many places, at so many times, and by such varied people, but the resulting concepts and laws transcend these varied origins.

It is nevertheless interesting that the discovery of energy was so Eurocentric. While the wheel, the windmill and the water-wheel, and so on cropped up all over the world, the steam engine, the electric motor, and the dynamo occurred in only one place, at one time. Doubtless there are geographical, socio-political, and economic factors, but the one-off nature of the discovery of these particular engines is ultimately a testimony to the rule of the Second Law: it is very difficult to get work from heat. This is evidenced by the extraordinary intricacy of these engines—see Watt's engine in Fig. 18.1 and Joule's engine in Fig. 14.1—so much more complicated than Hero's steam kettle, Pixii's dynamo, or dal Negro's motor.

One thing is sure, and that is that the concept of energy is here to stay. It is not sufficient for a concept just to be mathematically defined, measurable, and leading to consistent results—it must also get used. There is no doubt that 'energy' meets these requirements.

We have shown that the physics and the mathematics move forward together, but it is impossible to tell where the mathematics ends and the physics begins, and vice versa. This is the real message of Feynman's allegory and the reason why energy is such a slippery concept: the 'blocks' *are* the real thing, energy, and can be measured (in Joules); and they are also nothing more than the mathematical formulae $\frac{1}{2}mv^2$, $\int \mathbf{F}.d\mathbf{r}$, mgh, $\frac{1}{2}kT$, $\frac{1}{2}CV^2$, qV, mc^2, and so on.

This is not the place to address our current woes concerning diminishing resources and global warming. I will, nevertheless, offer two thoughts. One is that there is no safe form of energy—it is energetic, after all. The other is that, as the sink for all mankind's activities becomes warmer, so all our 'engines'—geophysical and biological as well as man-made—will work less and less efficiently. Cars already run less efficiently in the summer,[5] plants will have to consume even more water for cooling, the human 'engine' will work less well, and so on.

While the challenge has more to do with holding on to our humanity than with anything technological, it is nonetheless seductive to try and think of some 'hi-tech' solutions. One idea of mine was the use of synchronized rocket-firings to move the Earth into a new orbit, further away from the Sun. However, a back-of-the-envelope calculation soon showed that the amount of energy required would be astronomical, or, should I say, planetary. We must heed Pauli's advice, that in order to have one good idea, one must have many ideas. I hope this book will inspire many readers to have many ideas.

In one sentence, energy is: the ceaseless jiggling motion, the endless straining at the leash, even in apparently empty space, the rest mass and the radiation, the curvature of space–time, the foreground activity, the background hum, the *sine qua non*.

APPENDIX I

Timeline

[1618–48	The 'Thirty Years' War']
1620	Bacon, heat is a motion of the small parts in *Novum organum*
1620s–1630s	Galileo, 'v^2 proportional to h'; thermoscope; Galilean relativity of motion
1629–33	Descartes, 'The world'
1632	Galileo, 'Dialogue concerning the two chief systems of the world—Ptolemaic and Copernican'
[1634	Thomas Hobbes visits Galileo]
1637	Descartes, *Discours de la méthode*, including the *Dioptrique* Descartes (in letter to Huygens), 'force' is 'weight × height'
1638	Galileo, *Discourses Concerning the Two New Sciences* [John Milton visits Galileo]
1644	Torricelli's barometer and the 'ocean of atmosphere'
1646	Pascal's 'barometric' experiments with water and wine at Rouen
[1648	Thomas Hobbes meets Descartes]
1656	Huygens, *De motu corporum ex percussione*
1657	von Guericke's demonstration with evacuated copper hemispheres
1660	Boyle, *New Experiments Physico-mechanical, Touching the Spring of the Air, and its Effects*
[1660s–1670s	Benedict Spinoza]
1662	Boyle's Law and Fermat's Principle of least time (for the path of light)
1663	Marquis of Worcester, 'water-commanding' engine
1665	Boyle distinguishes motion of heat and bulk motion in *New Experiments and Observations Touching Cold*
[1665–66	The Great Plague of London, the fire of London]
1673	Huygens, *Horologium oscillatorium sive de motu pendularium*, dedicated to Louis XIV, and *De vi centrifuga*
1679	Mariotte, '*Essay du chaud et du froid*'; Papin's 'digester'
1686	Leibniz, *Brevis demonstration erroris memorabilis Cartesii* —he discovers mv^2 or *vis viva*

1687	Newton, *Principia*
[1690	John Locke, *An Essay Concerning Human Understanding*]
1692	Leibniz, *Essay de dynamique*
1695	Leibniz, *Specimen dynamicum*
1697	Johann Bernoulli's solution of the brachistochrone problem using the path of light through a medium of variable refractive index
1698	Savery's 'fire-engine' or 'miner's friend'
1699	Amontons's 'fire-mill' and definitions of work and friction
[*c*. 1700	Becher and Stahl, phlogiston theory]
1701	Newton's law of cooling (published anonymously)
1703	Jakob Bernoulli on the compound pendulum as a hypothetical lever
1704	Antoine Parent, 'Theory of the greatest possible perfection of machines'
1712	Newcomen's 'atmospheric' steam engine
1715	Brook Taylor, '*Methodus incrementorum directa et inversa*'
	Johann Bernoulli's 'Principle of virtual velocities/work' in a letter to Varignon; first use of word 'energy' in physics
1716	Jakob Hermann, first kinetic theory in *Phoromania*; Death of Leibniz
1720	's Gravesande, *Mathematical Elements of Natural Philosophy, Confirm'd by Experiments*
1720s	Experiments on balls falling into clay
	Fahrenheit carries out experiments on thermometers and on heat
1723	Brook Taylor uses the 'method of mixtures' to develop a scale of temperature
1724	*Vis viva* controversy continues with competition of the Académie Royale des Sciences on the 'communication of motion'
1727	Hales, *Vegetable Staticks*; Newton dies
1727–28	Johann Bernoulli's work on vibrating strings; Voltaire in exile in London
1733	Voltaire, *Philosophical Letters on the English*
	Daniel Bernoulli's work on vibrating strings and trigonometric series
1735	Boerhaave, *Elementa chemiae* and his subtle-fluid theory of heat
1736	Euler, '*Mechanica, sive motus scientia analytice exposita*', in which Newtonian mechanics is put in analytical form
1737	Algarotti, *Newtonianism for the Ladies*
	Emilie du Châtelet, '*Dissertation sur la nature et la propagation du feu*'
	Voltaire, 'Elements of the philosophy of Newton'

1738	Daniel Bernoulli, *Hydrodynamica*, including the kinetic theory of gases, the idea of *vis potentialis*, and the conservation of 'live force'
	Daniel Bernoulli's paper on the Sun–Earth–Moon system, demonstrating the route-independence of *vis viva* between fixed end-points
	Martine finds that 'quicksilver is [exceptionally] ticklish'
1740	Du Châtelet, *Lessons in Physics*
1741, 1744, and 1746	Maupertuis develops his 'Principle of Least Action'
1743	D'Alembert, *Traité de dynamique*
	Clairaut, *Théorie de la figure de la terre*, the start of potential function theory
[1748	David Hume, *An Enquiry Concerning Human Understanding*]
1749	du Châtelet translates Newton's *Principia* into French
1751–72	Diderot, *Encyclopédie* (some with d'Alembert)
1752	Déparcieux uses reversibility argument to quantify the efficiency of water-wheels
	[Voltaire, *The Diatribe of Dr Akakia, Citizen of St Malo*]
1756	Cullen, 'Of the cold produced by evaporating fluids and of some other means of producing cold'
1759	Smeaton, overshot water-wheels are more efficient than undershot wheels
1760s	Black develops his theories of latent and specific heat
1763	Boscovich, 'A theory of natural philosophy'
1765	Watt's idea of separate condenser
1769	Wilcke, latent and specific heat
1770s	Irvine's theory of heat
1773–80	Lagrange, intimations of potential function theory
1776	First commercial Boulton and Watt steam engine
	[Adam Smith, *The Wealth of Nations*]
	[American War of Independence]
1777	Scheele, 'A chemical treatise on air and fire', discovers radiant heat
1779	Crawford, first to consider the specific heat of gases in *Experiments and Observations on Animal Heat*
[*c*. 1780s on	The Industrial Revolution]
[1781	Immanuel Kant, *Critique of Pure Reason*]
	Hornblower's two-cylinder compound engine
1782	Watt's rotative engine
1783–84	Legendre functions (used by Laplace)
	Lavoisier and Laplace, *Memoir on Heat*, use of ice calorimeter

	Watt's first double-acting engine
	Lazare Carnot, '*Essai sur les machines en general*'
	[First flight in a hot-air balloon, made by the Montgolfier brothers]
1785–89	Ingen-Housz, '*Nouvelles éxperiences et observations sur divers objets de physique*', speed of heat-conductivity experiments
[1786	First ascent of Mont Blanc, by Balmat and Paccard]
[1786, 1787, and 1789	Mozart composes 'The Marriage of Figaro', 'Don Giovanni', and 'Cosi Fan Tutte']
1787?	Cavendish's kinetic theory in 'Heat', only discovered in 1969
[1788	Hutton, *Theory of the Earth*]
	Lagrange, *Analytique Mécanique*
1789	Laplace's theory of spheroidal attraction, applied to the rings of Saturn, leads to 'Laplace's equation'
	Lavoisier, *Traité élémentaire de chimie*, in which '*calorique*' is one of the elements
	[French Revolution]
	Rumford's generation of heat by boring cannons
1790	Pictet, '*Essai sur le feu*'
1791	Prévost, '*De l'equilibre du feu*'
1792	Davies Gilbert shows that $\int P\,dV$ is the work done by a steam engine
1794	Lavoisier loses his head at the guillotine
1796	Southern's 'indicator'
1799	Trevithick's first high-pressure steam engine
	Davy's ice-rubbing experiment
1799–1825	Laplace, *Mécanique Céleste*
1800	William Herschel discovers infrared radiation
	Volta invents the voltaic pile [Watt's patent expires]
1800s	Dalton, thermal expansion of gases, atomic theory
1801	Young puts forward the wave theory of light and also of heat
1802	Ritter discovers ultraviolet radiation
	Gay-Lussac, expansivity of gases with heat
1804	Fire-piston demonstration at Lyon
1806	Berthollet and Laplace found the Societé d'Arcueil
1807	Young, in his 'A course of lectures in natural philosophy and the mechanical arts', proposes the term 'energy' instead of 'living force'
	Gay-Lussac's law of simple proportions and the twin flasks experiment
[1810	Goethe, *Theory of Colours*]

1811	Avogadro's Hypothesis
1812	Clément and Desormes' measurement of γ
[1815	Battle of Waterloo]
1816	Laplace's 'adiabatic' correction to the speed of sound Herapath's kinetic theory of gases [Coleridge's poem 'Kubla Khan']
1819	Dulong and Petit's Law on the constancy of specific heat per atom
1820	Oersted, compass needle moves when near 'galvanic current'
1821–60	Faraday's experiments on the 'unity of force'
1821	Seebeck, link between electricity and heat
1821 (1822?)	Poisson's equation
1822	Fourier, *The Analytical Theory of Heat* [Charles Babbage starts making his first calculator, or 'difference engine']
1824	Sadi Carnot, *Réflexions sur la puissance motrice du feu* [First performance of Beethoven's Ninth Symphony]
[1827	'Congreves' friction matches]
1828	Green defines the potential function (in 'An essay on the mathematical analysis of electricity and magnetism')
[1829	Stephenson's 'Rocket'] Coriolis defines work as the integral of force over distance
1830	Hippolyte Pixii's dynamo
1831	dal Negro's electric motor
1833	Hamilton's mechanics
1834	Peltier, link between electricity and heat Lenz's Law; Clapeyron, 'Memoir on the motive power of heat'
1835	Ampère's wave theory of heat
[1837	Start of Queen Victoria's reign]
1841–48	Julius Robert Mayer's conservation of 'force' and mechanical equivalent of heat
[1842	Quetelet, *Treatise on Man*]
1843	James Joule, the mechanical equivalent of heat (from magneto-electricity) and the 'I^2R' law
1843–45	Waterston, kinetic theory of gases
1844–45	Joule's twin cylinder and paddle-wheel experiments
1846	Groves, 'On the correlation of physical forces'
1847	Joule's ideas noticed by William Thomson (Lord Kelvin) at the British Association meeting in Oxford; Helmholtz, '*Über der Erhaltung der Kraft*'
1850	Clausius, the First and Second Laws of Thermodynamics, in 'On the motive power of heat, and on the laws which can be deduced from it for the theory of heat'

1851	Thomson accepts dynamical theory of heat and conservation of energy; alternative statement of Second Law; introduces term 'energy'
1852	Thomson, 'On a universal tendency in nature to the dissipation of mechanical energy'
1854	Helmholtz, 'heat death' of the universe
1857	Clausius, kinetic theory of gases
[1859	Darwin's theory of evolution]
1865	Clausius coins term 'entropy' and states two principles: the energy in the universe is constant; the entropy of the universe tends to a maximum
1865–73	Maxwell's theory of electromagnetism
1867	Maxwell's kinetic theory of a gas
1872	Boltzmann factor, Maxwell–Boltzmann distribution
1875–78	Gibbs, 'On the equilibrium of heterogeneous substances'
1876	Boltzmann, equipartition theorem
1877	Boltzmann, microscopic formulation of entropy
1884	Poynting, defines flux of electromagnetic radiation
1886	Hertz discovers radio waves
1900	Planck, quantum of radiation energy, $E = h\nu$
1905	Einstein, Special Theory of Relativity, $E = mc^2$, light is particulate
1915	Einstein, General Theory of Relativity, leading later to the stress–energy tensor

APPENDIX II

Questions

There follows a list of questions of varying difficulty and in no particular order. About half are answered in the book.

(1) Leibniz took the variability in blades of grass as evidence that nature is not atomic. Why was he wrong?

(2) Carnot said that work can only be done when heat flows down a temperature gradient, yet in the isothermal sections of his Carnot cycle (Fig. 12.2), work is done and the temperature remains constant. Explain.

(3) Which crucial part of the system was not scaled up in going from the small bone to the giant's bone (see 'Galileo's bones', Fig. 3.1)?

(4) What is the difference between the path of a cannon ball dropped from the Leaning Tower of Pisa and one dropped from the crow's nest of a ship sailing (uniformly) from Venice to Aleppo?

(5) Could thermodynamics be recast with diabaric walls and isobars instead of diathermal walls and isotherms?

(6) Is Newton's Third Law of Motion ever wrong?

(7) Are there just two main kinds of energy, kinetic and potential, and is one more fundamental than the other?

(8) How is Lenz's Law explained at the microscopic level?

(9) In the free expansion of a gas, how is it that the temperature doesn't drop, given that it increases when the gas is compressed back to its original volume?

(10) Apart from the free expansion of a gas, are there any other processes in nature in which the entropy increases but the temperature stays constant, no heat is added and no work is done?

(11) How can energy be conserved for a central force that is not of the 'inverse-square' form?

(12) In changing between two frames of reference in uniform relative motion, where does the kinetic energy disappear to?

(13) How can there be a cosmic arrow of time when, on the microscopic scale, time is reversible?

(14) Extensive quantities, such as mass, volume, energy, and entropy, change after a change of scale. What happens to intensive quantities, such as density or temperature?

(15) What are the three kinds of perpetual motion discussed in this book?
(16) How can Boltzmann's microscopic formulation of entropy be identified with Clausius's entropy?
(17) How do 'engines' and 'work' come into physics?
(18) Why is the story of energy so Eurocentric?
(19) Is 'mc^2' a more fundamental form of energy than the other blocks?
(20) What is the answer to Galileo's headache about dead weights (see the end of the Galileo section in Chapter 3)?
(21) What happens to the entropy of a black hole as mass-energy gets swallowed up?
(22) Does the present matter phase of the universe (with stars, galaxies, etc.) represent a higher or lower entropy than the radiation phase of the universe? If lower, then how does this square with the Second Law?
(23) Does Darwin's theory of evolution contradict the Second Law?
(24) Everything we experience is at dis-equilibrium (the universe is expanding and we're on a warm planet in cold space and exposed to the Sun's hot rays). How can we develop a physics based on equilibrium states?
(25) What is 'action' and why is it important in physics?
(26) Why is the Lagrangian given by $(T - V)$? (T is kinetic energy, V is potential energy.)
(27) Which is more important, $(T + V)$ or $(T - V)$?
(28) Is the Lagrangian sometimes maximized rather than minimized and how would this affect the arguments of Chapter 7, part IV?
(29) Why does one and the same T (temperature) apply in classical thermodynamics, statistical mechanics, and radiation theory?
(30) Can it ever be meaningful to talk of the 'heat-in-a-body'?
(31) Is there any friction between the tyres of a car and the road surface (assuming no skids)?
(32) Are the laws of thermodynamics empirical or necessary?
(33) Can the universe be regarded as one thermodynamic system?
(34) What is the evidence that 'heat is the motion of the small parts'?
(35) Will my can of fizzy drink cool down if I open it quickly and let the fizz rush out?
(36) How do directions get randomized in the elastic collisions in an ideal gas?
(37) How does the bulk forward motion of a piston in a gas get 'lost', and how is the total momentum of the system conserved?
(38) Explain qualitatively why a gas doesn't become less viscous as its temperature is increased.
(39) What is the connection between 'energy' and 'time'?
(40) Is 'force' or 'energy' more important/elemental in physics?
(41) How do the scalar quantities T and V (kinetic and potential energy) take account of the directionality of motion?

(42) What are the links between the measures of electrical energy in circuits (VIt and $\frac{1}{2}CV^2$), the potential function $V(r)$, and the energy of the electric field?
(43) Can ideally hard billiard balls undergo elastic collisions?
(44) Can the Second Law of Thermodynamics be sloganized as 'nature abhors a gradient'?
(45) How crucial was Watt's steam engine (the separate condenser, the 'expansive operation', and the 'indicator diagram') to the development of thermodynamics?
(46) In an endless succession of complete Carnot cycles (forward direction), total entropy and total energy are conserved and yet more and more work gets done. Explain.
(47) Draw the Carnot cycle (the PV diagram) for Clausius's most general case of heat transferred between two temperatures and heat converted into work at a third temperature.
(48) Instead of using lots of pre-scaled ideal heat-engines to set up the absolute temperature scale (see the beginning of Chapter 16), we could just as well employ one ideal heat engine and let it chart the decreasing temperature each time one unit of work is put out. Explain how the two temperature scales can be made equivalent.
(49) Evans and Popp* describe an experiment where it appeared that 'cold' *was* transmitted and reflected. Think about how this could have been achieved.
(50) Is the magnetism in loadstone or a bar magnet a 'relativistic effect'?
(51) A ball bounces elastically from a wall. How can it be ensured that total momentum is conserved, and without sacrificing energy conservation?
(52) In the case of elastic two-body collisions with the target particle at rest, the kinetic energy per particle is the same or closer to the average *after* the collision. Explain why this is not the origin of time's cosmic arrow.

* See Evans, James and Popp, Brian, Pictet's experiment, *American Journal of Physics*, 53(8), 1985, pp. 737–53.

Bibliography

There is nothing more enlightening than going back to the primary sources (as given in the Notes and References section). For secondary material, six marvellous books have been my guide at every stage (see the Acknowledgments). The following books have also been invaluable as background for individual chapters:

Atkins, Peter, *The Second Law: Energy, Chaos and Form*, Scientific American Library, 1994.

Brush, Stephen G., *The Temperature of History*, Lennox Hill, 1979.

Cardwell, Donald, *James Joule, a Biography*, Manchester University Press, 1989.

Drake, Stillman, *Galileo*, Oxford University Press, 1980.

Fox, Robert, *The Caloric Theory of Gases from Lavoisier to Regnault*, Clarendon Press, 1971.

Gillispie, Charles C., *Lazare Carnot, Savant*, Princeton University Press, 1971.

Grattan-Guiness, Ivor, *Convolutions in French Mathematics: 1800–1840*, Birkhauser, 1990.

Guerlac, Henry, *Essays and Papers in the History of Modern Science*, Johns Hopkins University Press, 1977.

Hankins, Thomas, *Sir William Rowan Hamilton*, Johns Hopkins University Press, 1977.

Heilbron, John L., *Electricity in the Seventeenth and Eighteenth Centuries: a Study of Early Modern Physics*, University of California Press, 1979.

Jungnickel, Christa and McCormmach, Russell, *Cavendish, The Experimental Life*, Bucknell, 1999.

Klein, Martin, Thermodynamics in Einstein's Thought, *Science*, 157, 1967, pp. 509–16.

Knowles Middleton, W.E., *A History of the Thermometer and its Uses in Meteorology*, Johns Hopkins University Press, 1966.

Müller, Ingo, *A History of Thermodynamics: the Doctrine of Energy and Entropy*, Springer-Verlag, 2007.

Pearce Williams, L., *Michael Faraday, a Biography*, Da Capo Press, 1987.

Smith, Crosbie, *The Science of Energy: A Cultural History of Energy Physics in Victorian Britain*, The Athlone Press, 1998.

Smith, Crosbie, and Wise, Norman M., *Energy and Empire: a Biographical Study of Lord Kelvin*, Cambridge University Press, 1989.

Truesdell, Clifford, The rational mechanics of flexible or elastic bodies, 1638–1788, introduction to *Leonhardi Euleri Opera Omnia*, 2nd Series, Fussli, 1960.

Westfall, R., *The Construction of Modern Science*, Cambridge University Press, 1977.

Notes and References

Two sources will be abbreviated as follows:

DSB = *Dictionary of Scientific Biography*, editor-in-chief Charles C. Gillispie, Charles Scribner's Sons, New York (1970–80) and since December 2007 available as an e-book.

Cardwell, W to C = Cardwell, Donald S.L., *From Watt to Clausius: the Rise of Thermodynamics in the Early Industrial Age*, Heinemann, London, 1971.

CHAPTER 1, PAGES 3–4

1. Lanczos, Cornelius, *The Variational Principles of Mechanics*, 4th edn, University of Toronto Press, Toronto, 1974, Preface, p. x.
2. Feynman, Richard, *Lectures on Physics*, with R. Leighton and M. Sands, vol. 1, Addison-Wesley, Reading, MA, 1966.
3. Ibid.

CHAPTER 2, PAGES 5–12

1. Coverage based on Hiebert, Erwin N., *Historical Roots of the Principle of the Conservation of Energy*, Arno Press, New York, 1981; and Ord-Hume, Arthur W.J.G., *Perpetual Motion: the History of an Obsession*, St Martin's Press, New York, 1980.
2. Aristotle's *Mechanica* 1–3, as found in Hiebert, *Historical Roots*, ch. 1, this and following quotes.
3. Needham, Joseph, *Science in Traditional China: a Comparative Perspective*, Harvard University Press, Cambridge, MA, 1981.
4. Hero's *Mechanica*, as referenced and found in Hiebert, *Historical Roots*, ch. 1.
5. Sarma, S.R., 'Astronomical instruments in Brahmagupta's Brahmasphutasiddhanta', *Indian Historical Review*, XII, 1986–7, p. 69, as found at www.hp-gramatke.de.net/perpetuum/english/page0220.htm
6. Ord-Hume, *Perpetual Motion*, ch. 4, p. 63.
7. Somerset, Edward, Marquis of Worcester, 'A century of inventions, written in 1655', printed by J. Grismond, 1663; electronic version at www.history.rochester.edu/steam/dircks/ (with thanks to Fran Versace and Sean Singh).
8. Tallmadge, G.K., 'Perpetual motion machines of Mark Antony Zimara', *Isis*, 33(8), 1941, pp. 8–14.
9. Wilkins, John, 'Mathematical magick', London, 1648 (as given in Ord-Hume, *Perpetual Motion*).

10. Stevin, Simon, *De Beghinselen der Weeghconst* (*The Laws of Statics*), 1586; and Devreese, J.T. and Vanden Berghe, G., *'Magic is No Magic': the Wonderful World of Simon Stevin*, WIT Press, Southampton, 2008.

CHAPTER 3, PAGES 14–45

1. Gillispie, Charles C., *The Edge of Objectivity*, Princeton University Press, Princeton, NJ, 1973.
2. I owe this deep insight to Charles Gillispie, *The Edge of Objectivity*.
3. Galileo, *Two New Sciences, Including Centres of Gravity and Forces of Percussion* (1638), translated with introduction and notes by Stillman Drake, University of Wisconsin Press, Madison, WI, 1974, p. 166.
4. Galileo, *Dialogue Concerning the Two Chief World Systems—Ptolemaic and Copernican*, 2nd rev. edn, translated with notes by Stillman Drake, foreword by Albert Einstein, University of California Press, Berkeley, CA, 1967, p. 239.
5. Drake, Stillman, 'Galileo's discovery of the law of free fall', *Scientific American*, 228(5), 1973, p. 84.
6. Galileo, *Two New Sciences*, p. 68.
7. Drake, Stillman, *Galileo*, Oxford University Press, Oxford, 1980; and Koestler, Arthur, *The Sleepwalkers*, Penguin, London, 1986. (Koestler, in his otherwise excellent book, has Kepler as the favourite and Galileo as the un-favourite.)
8. Drake, Stillman, *Discoveries and Opinions of Galileo*, Doubleday, New York, 1957.
9. Galileo, *Two Chief World Systems*, p. 28.
10. Galileo, *Two Chief World Systems*, p. 147.
11. Galileo, *Two Chief World Systems*, p. 116.
12. Galileo, *Two Chief World Systems*, pp. 186–7.
13. Galileo, *Two Chief World Systems*, the postil on p. 273.
14. Drake, Stillman, *Galileo at Work*, University of Chicago Press, Chicago, 1978, p. 298.
15. Galileo, *Dialogues Concerning Two New Sciences* (1638), translated by H. Crew and A. de Salvio, edited by A. Favaro, Dover Publications, New York, 1954; reprint of 1914 edition, 'Fourth Day', p. 293.
16. Ibid., p. 271.
17. Sorrel, Tom, *Descartes*, Past Masters, Oxford University Press, Oxford, 1987.
18. Descartes, René, 'The World' in *Discourse on Method and Related Writings*, 1637; Penguin Classics edition, London, 1999, translated and introduced by Desmond Clarke.
19. Descartes, René, *The Principles of Philosophy* (1644), Part II, in Haldane, E.S. and Ross, G.R.T., *The Philosophical Works of Descartes*, Dover Publications, New York, 1955.

20. Descartes, René, 'The world', in *Discourse on Method*.
21. Descartes, René, *The Principles of Philosophy*.
22. Westfall, Richard, *Force in Newton's Physics*, American Elsevier and Macdonald, London, 1971, ch. II, this and the following quotes in Descartes' letters to Mersenne and Huygens.
23. Descartes, René, letter to Mersenne, July 1638, in *Oeuvres de Descartes*, vol. II ('*ma physique n'est autre chose que géométrie*').
24. Huygens, Christiaan, *Horologium oscillatorium*, 1673; original Latin on 'Gallica' website of the French National Library; English translations by Richard Blackwell (1986) and Ian Bruce (2007).
25. Huygens, Christiaan, *De motu de corporum ex percussione*, 1656, in *Oeuvres complète*, Martinus Nijhoff, La Haye, 1888–1950, vol. 16.
26. Bell, A.E., *Christian Huygens and the Development of Science in the Seventeenth Century*, Longmans, Green & Co., New York, 1947.
27. Huygens, *De motu*. (I have translated the hypotheses and propositions from seventeenth-century French, itself a translation from Huygens' Latin.)
28. Gabbey, Alan, 'Huygens and mechanics', in Bos, H.J.M., Rudwick, M.J.S., Snelders, H.A.M., and Visser, R.P.W. (eds), *Studies on Christiaan Huygens, Symposium on 'The Life and Works of Christiaan Huygens', Amsterdam, 1979*, Swets and Zeitlinger, Lisse, 1980.
29. Huygens, *De motu*.
30. Westfall, *Force in Newton's Physics*, ch. IV.
31. Huygens, *De motu*.
32. Gabbey, 'Huygens and mechanics'.
33. Crew, H., in Bell, *Christian Huygens*.
34. Newton, Isaac, *The Mathematical Principles of Natural Philosophy* (1687), translated 1729 by Andrew Motte, Daniel Adee, New York, 1848; and in Hawking, Stephen (ed.), *On the Shoulders of Giants*, Running Press, Philadelphia, PA, 2002.
35. Westfall, Richard, *Never at Rest: a Biography of Isaac Newton*, Cambridge University Press, Cambridge, 1983.
36. Newton, *The Mathematical Principles of Natural Philosophy*.
37. Ibid.
38. Ibid.
39. Ibid., paragraph following Law III.
40. Ibid., Scholium after corollary VI.
41. Maxwell, James Clerk, *Matter and Motion*, Dover Publications, New York, 1952.
42. Newton, *The Mathematical Principles of Natural Philosophy*, Scholium after corollary VI.
43. Ibid.
44. Home, Roderick W., 'The Third Law in Newton's mechanics', *British Journal for the History of Science*, 4(13), 1968, p. 39.

45. Newton, Isaac, 'Queries' at the end of *Opticks: or, a Treatise of the Reflections, Refractions, Inflections and Colours of Light*, 2nd edn, 1717, translated by Andrew Motte, revised by Florian Cajori, as in Great Books of the Western World, no. 32, Encyclopaedia Britannica, Inc., University of Chicago, 2nd edn, 1990.
46. Newton, Isaac, 'Queries', Query 31, p. 541.
47. Ibid., Query 31 pp. 540–1.
48. At least, I couldn't find anything further in Westfall, *Never at Rest*.
49. Newton, *The Mathematical Principles of Natural Philosophy*; see all of Book II.
50. Newton, 'Queries', Query 28, p. 528.
51. Ibid., Query 31, p. 540.
52. Ibid., Query 31, p. 542.
53. MacDonald Ross, G., *Leibniz*, Past Masters, Oxford University Press, Oxford, 1984.
54. Leibniz, Gottfried Wilhelm, *Philosophical Writings*, edited by G.H.R. Parkinson, Dent, London, 1973.
55. Ibid.
56. Ibid.
57. Leibniz, Gottfried Wilhelm, 'Brief demonstration…' (1686), in *Philosophical Papers and Letters*, translated and edited by Leroy Loemker, Chicago University Press, Chicago, 1956.
58. Leibniz, Gottfried Wilhelm, 'Essay on dynamics', in Costabel, Pierre, *Leibniz and Dynamics; the Texts of 1692*, Hermann, Paris/Cornell University Press, Ithaca, New York, 1973.
59. Westfall, *Force in Newton's Physics*, p. 297.
60. Ibid., p. 297.
61. Ibid., discussion on pp. 297–303.
62. DSB on Bernoulli, Johann.
63. Westfall, *Force in Newton's Physics*, p. 288.
64. Westfall, *Force in Newton's Physics*.
65. Leibniz, *Philosophical Writings*, p. 158.
66. Westfall, *Force in Newton's Physics*, p. 295.

CHAPTER 4, PAGES 46–61

1. Bacon, Francis, *The New Organon*, 1620; translated in 1863 by Spedding, Ellis and Denon Heath, The Aphorisms, Book Two, XX.
2. Gassendi, Pierre, *Epicuri philosophiae syntagma*, 1649.
3. Boyle, Robert; this and following quotes of Boyle excerpted from Boas Hall, Marie, *Robert Boyle on Natural Philosophy*, Indiana University Press, Bloomington, IN, 1965; and Boas, Marie, *Robert Boyle and Seventeenth Century Chemistry*, Cambridge University Press, Cambridge, 1958.
4. Maxwell, James Clerk, *Theory of Heat*, 1871; reissued Dover Publications, New York, 2001, with a new introduction and notes by Peter Pesic; see p. 303.

5. Boyle, Robert, *New Experiments and Considerations Touching Cold*, printed for J. Crook, London, 1665.
6. Hooke, Robert, *Micrographia*, 1665, p. 46 of original; see also Dover Phoenix Editions, New York, 2003.
7. Cardwell, W to C, ch. 1.
8. Knowles Middleton, W.E., *A History of the Thermometer and its Uses in Meteorology*, Johns Hopkins University Press, Baltimore, MD, 1966, p. 8.
9. Knowles Middleton, *A History of the Thermometer*, p. 27.
10. Boyle, Robert, in Boas Hall, *Robert Boyle on Natural Philosophy*.
11. Ibid.
12. Galileo, *Two New Sciences, Including Centres of Gravity and Forces of Percussion* (1638), translated with introduction and notes by Stillman Drake, University of Wisconsin Press, Madison, WI, 1974, 'First Day'.
13. Knowles Middleton, W.E., *The History of the Barometer*, Johns Hopkins University Press, Baltimore, MD, 1964, p. 5.
14. Torricelli, Evangelista, 'we live submerged at the bottom of an ocean of air' (1644), in a letter. See also note 13.
15. Boyle, Robert, *New Experiments Physico-mechanical, Touching the Spring of the Air, and its Effects*, 1662; second edition, 'Whereunto is added a defence of the author's explication of the experiments, against the objections of Franciscus Linus, and, Thomas Hobbes', printed by H. Hall for Tho. Robson, Oxford.
16. Ibid.
17. Mariotte, Edme, '*Essai du chaud et du froid*' (1679), in *Oeuvres de Mariotte*, vol. I, P. van der Aa, Leiden, 1717, p. 183.
18. Knowles Middleton, W.E., *The Experimenters, a Study of the Accademia del Cimiento*, Johns Hopkins University Press, Baltimore, MD, 1971.
19. Cardwell, W to C, ch. 1.
20. Papin, Denis, *A Continuation of the New Digester of Bones*, London, 1687.
21. Amontons, Guillaume, 'Moyen de substituer commodement l'action du feu…', *Mémoires de l'Academie des Sciences*, 1699, p. 112.
22. Amontons, Guillaume, 'De la résistance causée dans les machines…', *Mémoires de l'Academie des Sciences*, 1699, p. 206.

CHAPTER 5, PAGES 64–77

1. Boyle, Robert, *New Experiments Physico-Mechanical, Touching the Spring of the Air, and its Effects*, 1662, printed by H. Hall for Tho. Robson, Oxford.
2. Newton, Isaac, *The Mathematical Principles of Natural Philosophy* (1687), translated by Andrew Motte, revised by Florian Cajori, as in Great Books of the Western World, no. 32, Encyclopaedia Britannica, Inc., University of Chicago, 2nd edn, 1990, Book II, Proposition 23.

3. Newton, Isaac, 'Queries' at the end of *Opticks: or, a Treatise of the Reflections, Refractions, Inflections and Colours of Light*, 2nd edn, 1717, translated by Andrew Motte, revised by Florian Cajori, as in Great Books of the Western World, no. 32, Encyclopaedia Britannica, Inc., University of Chicago, 2nd edn, 1990.
4. Newton, Isaac, 'A scale of the degrees of heat', *Philosophical Transactions of the Royal Society*, 1701, p. 824.
5. Hales, Stephen, *Vegetable Staticks*, Experiment LXVII, ch. VI, p. 101, first published (1727) and published by The Scientific Book Guild, Oldbourne Book Co Ltd (1961), foreword by M.A. Hoskin.
6. Leicester, Henry M., *The Historical Background to Chemistry*, Dover Publications, New York, 1956.
7. Hales, *Vegetable Staticks*.
8. Boerhaave, Hermann, *The Elements of Chemistry*, translated from the original Latin by Timothy Dallowe, London; printed for J. & J. Pemberton, Fleet Street; J. Clarke, Royal Exchange; A. Millar in the Strand and J. Gray in the Poultry, London, 1735.
9. Boerhaave, Hermann; this and subsequent quotes are all taken from *The Elements of Chemistry*.
10. DSB on Brook Taylor.
11. Boerhaave, *The Elements of Chemistry*; and Fahrenheit, Daniel Gabriel, in Van der Star, P. (ed.), *Fahrenheit's Letters to Leibniz and Boerhaave*, Amsterdam, Rodopi, 1983.
12. Martine, George, *Essays and Observations on the Constitution and Graduation of Thermometers and on the Heating and Cooling of Bodies*, 3rd edn, Edinburgh, 1780.
13. Hermann, Jakob, *Phoromania*, Amsterdam, 1716.
14. Knowles-Middleton, W.E., 'Jacob Hermann and the kinetic theory', *British Journal for the History of Science*, 2, 1965, pp. 247–50.
15. Ibid.
16. Bernoulli, Daniel, *Hydrodynamica* (*Hydrodynamics*), 1738; reissued by Dover Publications, New York, 1968; all subsequent quotations in this section relate to ch. X.
17. Bernoulli, *Hydrodynamica*, ch. X.
18. Boerhaave, *The Elements of Chemistry*.
19. Cardwell, Donald S.L., *Technology, Science and History*, Heinemann Educational, London, 1972.
20. Cardwell, *Technology, Science and History*.

CHAPTER 6, PAGES 79–89

1. Uglow, Jenny, *The Lunar Men: the Friends Who Made the Future, 1730–1810*, Faber and Faber, London, 2002.
2. Ross, I.S., *The Life of Adam Smith*, Oxford Scholarship Online, 1995.

3. Guerlac, Henry, *Essays and Papers in the History of Modern Science*, Johns Hopkins University Press, Baltimore, MD, 1977, ch. 18.
4. Guerlac, *Essays and Papers*; and also in Magie, W.F., *A Source Book in Physics*, McGraw-Hill, New York, 1935; and Black, Joseph, *Lectures on the Elements of Chemistry*, John Robison, Edinburgh, 1803.
5. Ibid.
6. Ibid, this and following quotes until note 7.
7. Heilbron, J.L., *Electricity in the Seventeenth and Eighteenth Centuries: a Study of Early Modern Physics*, University of California Press, Berkeley, CA, 1979, p. 85.
8. Guerlac, *Essays and Papers*; Magie, *A Source Book in Physics*.
9. Ibid.
10. Ibid.
11. Martine, George, *Essays and Observations on the Constitution and Graduation of Thermometers and on the Heating and Cooling of Bodies*, 3rd edn, Edinburgh, 1780.
12. Guerlac, *Essays and Papers*.
13. Heilbron, *Electricity in the Seventeenth and Eighteenth Centuries*; and in McKie, D. and Heathcote, N.H. de V., *The Discovery of Specific and Latent Heat*, Edward Arnold, London, 1935.
14. Crawford, Adair, *Experiments and Observations on Animal Heat and the Inflammation of Combustible Bodies*, John Murray, London, 1779; this and subsequent quotes taken from the second edition, 1788.
15. Cardwell, W to C, ch. 4, p. 90.
16. Heilbron, *Electricity in the Seventeenth and Eighteenth Centuries*.
17. Scheele, Carl Wilhelm, 'A chemical treatise on air and fire' (1777), in *Collected Papers of Carl Wilhelm Scheele*, translated by L. Dobbin, G. Bell & Sons, London, 1931; this and following quotes.
18. Cardwell, W to C, ch. 4.
19. Pictet, Marc-Auguste, *Essai sur le feu*, Geneva, 1790, and English translation by W. Belcombe, London, 1791.
20. Ibid.
21. Evans, James and Popp, Brian, Pictet's experiment, *American Journal of Physics*, 53(8), 1985, pp. 737–53.
22. Prévost, Pierre, *Du calorique rayonnant*, Paris and Geneva, 1809; and 'Mémoire sur l'equilibre du feu', *Journal de Physique (Paris)*, 28, 1791, pp. 314–22.

CHAPTER 7, PAGES 91–140

1. Newton, Isaac, *The Mathematical Principles of Natural Philosophy* (1687), translated 1729 by Andrew Motte, Daniel Adee, New York, 1848; and in Hawking, Stephen (ed.), *On the Shoulders of Giants*, Running Press, Philadelphia, PA, 2002.

2. Hankins, Thomas, 'Eighteenth century attempts to resolve the *vis viva* controversy', *Isis*, 56(3), 1965, pp. 281–97.
3. Ibid.
4. Laudan, L.L., 'The *vis viva* controversy, a post-mortem', *Isis*, 59, 1968, p. 139.
5. Westfall, Richard, *Never at Rest: a Biography of Isaac Newton*, Cambridge University Press, Cambridge, 1983.
6. Hankins, 'Eighteenth century attempts to resolve the *vis viva* controversy'.
7. Newton, Isaac, 'Queries' at the end of *Opticks: or, a Treatise of the Reflections, Refractions, Inflections and Colours of Light*, 2nd edn, 1717, translated by Andrew Motte, revised by Florian Cajori, as in Great Books of the Western World, no. 32, Encyclopaedia Britannica, Inc., University of Chicago, 2nd edn, 1990, Query 31.
8. DSB on Bernoulli, Johann.
9. Newton, 'Queries', Query 31.
10. Newton, *The Mathematical Principles of Natural Philosophy*, Proposition XXXIX of Book I.
11. Parent, Antoine, 'Theory of the greatest possible perfection of machines', 1704.
12. Westfall, Richard, *Force in Newton's Physics*, American Elsevier and Macdonald, London, 1971; and Parkinson, G.H.R. (ed.), *Philosophical Writings*, Dent, London, 1973.
13. Ibid.
14. Ibid.
15. Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972.
16. Clairaut, Alexis-Claude, *Théorie de la figure de la terre tirée des principes dé l'hydrostatique*, Paris, 1743; translated into German, Italian, and Russian.
17. Bernoulli, Daniel, *Hydrodynamica* (*Hydrodynamics*) (1738), translated by T. Carmody and H. Kobus, preface by Hunter Rouse, Dover Publications, New York, 1968.
18. Ibid.
19. Ibid.
20. Euler, Leonhard, *Mechanica, sive motus scientia analytice exposita*, 2 vols, St Petersburg, 1736; in *Opera Omnia Leonhardi Euleri*, 2nd series, I and II, (1911 onwards) Berlin etc.
21. Euler, Leonhard, *Letters to a German Princess*, written between 1760 and 1762), translated by Henry Hunter, 1795 edition, notes by David Brewster, vol. 1; reprinted by Thoemmes Press, Bristol, 1997.
22. Heilbron, J.L., *Electricity in the Seventeenth and Eighteenth Centuries: a Study of Early Modern Physics*, University of California Press, Berkeley, CA, 1979, p. 45.

23. Boscovich, R.G., *A Theory of Natural Philosophy*, 1763; translated by J.M. Child, The MIT Press, Cambridge, MA, 1966.
24. Darwin, Erasmus, 'The temple of nature' and 'The loves of the plants', published together as *The Botanic Garden* (1803); reissued by Scholar Press, Menston, 1973.
25. Bernoulli, *Hydrodynamica*.
26. Ibid.
27. Ibid.
28. Ibid, this and the other quotes in the paragraph.
29. Ibid.
30. Ibid.
31. Ibid.
32. Ibid.
33. Ibid.
34. Hales, Stephen, *Vegetable Staticks*, 1727, Experiment LXVII, ch. VI, p. 101; reissued by The Scientific Book Guild, Oldbourne Book Co. Ltd, 1961, with a foreword by M.A. Hoskin.
35. Bernoulli, *Hydrodynamica*.
36. Ibid.
37. Ibid.
38. Truesdell, Clifford, 'The rational mechanics of flexible or elastic bodies, 1638–1788', introduction to *Leonhardi Euleri Opera Omnia*, 2nd series, vols X and XI, Fussli, pp. 173–4.
39. Ibid., p. 174.
40. Ibid., p. 193.
41. Ibid., p. 199.
42. Euler, Leonhard, '*Additamentum I de curvis elasticis*', added to *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes*, Lausanne and Geneva, 1744; in *Leonhardi Euleri Opera Omnia*, 1st series, vol. XXIV, pp. 231–97; and the English translation in Oldfather, W.A., Ellis, C.A., and Brown, D.M., 'Leonhard Euler's elastic curves', *Isis*, 20, 1933, pp. 72–160.
43. Bernoulli, *Hydrodynamica*.
44. Euler, Leonhard, letter to Daniel Bernoulli from St Petersburg, 13 September 1738, in *Die Werke von Daniel Bernoulli*, Band 3, *Mechanik*, general editor David Spieser, Birkhäuser Verlag, Basel, 1987, p. 72; translated from German by Sabine Wilkens.
45. Bernoulli, *Hydrodynamica*, preface to the English translation, p. xi.
46. Dugas, René, *A History of Mechanics*, Dover Publications, New York, 1988, foreword by Louis de Broglie, translated by J.R. Maddox, p. 246; quotations of d'Alembert from his *Traité de Dynamique* (1743), edition of 1758.
47. Kline, *Mathematical Thought*.

48. Maupertuis, Pierre-Louis M de, '*Loi du repos des corps*' (1741), '*Accord de différentes loix de la nature qui avoient jusqu'ici paru incompatibles*' (1744), and '*Les loix du movement et du repos déduites d'un principe metaphysique*' (1746).
49. Euler, '*Additamentum I de curvis elasticis*'.
50. Dugas, *A History of Mechanics*, p. 273.
51. DSB on Maupertuis.
52. Voltaire, F.-M.A. de, *Histoire du Docteur Akakia et du Natif de St Malo* (*The Diatribe of Dr Akakia, Citizen of St Malo*), 1752.
53. Euler, Leonhard, p. 19 of 'Life of Euler' in *Letters to a German Princess*, letters between (1760–2), translated by Henry Hunter (1795), notes by David Brewster, reprinted by Thoemmes Press, Bristol, 1997.
54. DSB on Lagrange, p. 569.
55. Bernoulli, Johann, 26 February 1715, published posthumously in Pierre Varignon's *Nouvelle Mécanique* (1725).
56. Dugas, *A History of Mechanics*, discussion following Section 2, p. 244.
57. Dugas, *A History of Mechanics*, p. 246; and in d'Alembert's *Traité de Dynamique*.
58. Truesdell, Clifford, 'Rational mechanics', p. 186.
59. DSB on d'Alembert.
60. Ibid.
61. Dugas, *A History of Mechanics*, p. 244.
62. Lagrange, J.L., *Analytique Mécanique* (*Analytical Mechanics*) (1788) translated and edited by Auguste Boissonnade and Victor Vagliente, Kluwer Academic, Boston, 1997.
63. DSB on Lagrange, p. 561.
64. Ibid.
65. I owe most of this analysis to Feynman's *Lectures on Physics*, vol. II, ch. 19.

CHAPTER 8, PAGES 148–165

1. Henry Guerlac, 'Some aspects of science during the French Revolution', in *Essays and Papers in the History of Modern Science*, Johns Hopkins University Press, Baltimore, MD, 1977.
2. Marsden, Ben, *Watt's Perfect Engine*, Icon Books, Cambridge, 2002, p. 90.
3. Bernal, J.D., *Science in History*, Penguin, London, 1969.
4. Uglow, Jenny, *The Lunar Society*, Faber and Faber, London, 2002.
5. Robison, John, letter 213, Robison to Watt, in E. Robinson and D. McKie (eds), *Partners in Science, Letters of James Watt & Joseph Black*, Harvard University Press, Cambridge, MA, 1969.
6. Marsden, *Watt's Perfect Engine*, p. 58.
7. Cardwell, W to C, pp. 42–5.
8. Marsden, *Watt's Perfect Engine*, p. 132.
9. Cardwell, W to C, p. 80.

10. Marsden, *Watt's Perfect Engine*, p. 119.
11. DSB article on Watt.
12. Ibid.
13. Cardwell, W to C, p. 79.
14. Bronowski, J., *The Ascent of Man*, BBC Publications, London, 1973.
15. Bernal, *Science in History*.
16. Doherty, Howard, of Doherty's Garage, St Andrews Street, Bendigo, personal communication.
17. Carnot, Lazare, *Essai sur les machines en général*, 1783.
18. Gillispie, Charles C., *Lazare Carnot, Savant*, Princeton University Press, Princeton, NJ, 1971.
19. DSB on Lazare Carnot, p. 78.
20. Gillispie, Charles C., *The Edge of Objectivity: an Essay in the History of Scientific Ideas*, Oxford University Press, London, 1960, p. 205.
21. Lavoisier, Antoine-Laurent, *Memoir sur la Chaleur*, presented in 1783 and published in 1784; translated by Henry Guerlac.
22. Mendoza, E., Introduction to Sadi Carnot's *Reflections on the Motive Power of Fire*, Dover Publications, New York, 1960, p. xv.
23. Biographical notes to Lavoisier's *Elements of Chemistry*, Robert Kerr's translation, Britannica Great Books, vol. 45, 1952.
24. Ibid.
25. Guerlac, Henry, *Antoine-Laurent Lavoisier, Chemist and Revolutionary*, Charles Scribner's Sons, New York, 1973.
26. Grattan-Guiness, Ivor, in Gillispie, Charles C., *Pierre-Simon Laplace, 1749–1829*, Princeton University Press, Princeton, NJ, 2000.
27. Gillispie, Charles C., Laplace's biographer, personal communication, 2007.
28. Lavoisier, *Memoir sur la Chaleur*, this and subsequent quotes.
29. Gough, J.B., 'The origins of Lavoisier's theory of the gaseous state', in Woolf, H. (ed.), *The Analytic Spirit, Essays in the History of Science, in Honour of Henry Guerlac*, Cornell University Press, Ithaca, NY, 1981.
30. Lavoisier, Antoine-Laurent, *The Elements of Chemistry*, 1783, Second Part; in Britannica Great Books, vol. 45, 29th edn (1987), this and subsequent quotes.
31. Guerlac, Henry, *Antoine-Laurent Lavoisier, Chemist and Revolutionary*, Charles Scribner's Sons, New York, 1973.
32. Jungnickel, Christa and McCormmach, Russell, *Cavendish, The Experimental Life*, Bucknell, Lewsiburg, PA, 1999, p. 491.
33. Ibid., pp. 304–5.
34. McCormmach, Russell, 'Henry Cavendish on the theory of heat', *Isis*, 79(1), 1988, pp. 37–67.
35. McCormmach, 'Henry Cavendish on the theory of heat', note 4 on p. 39; Cavendish, Henry, 'Obs on Mr Hutchins's expts for determining the degree

of cold at which quicksilver freezes', *Philosophical Transactions of the Royal Society*, 73, 1783, pp. 303–28.

36. McCormmach, Russell, *The Speculative Truth: Henry Cavendish, Natural Philosophy and the Rise of Modern Theoretical Science*, Oxford University Press, Oxford, 2004.
37. Ibid.
38. Jungnickel and McCormmach, *Cavendish, the Experimental Life*, p. 409.
39. McCormmach, *The Speculative Truth*.

CHAPTER 9, PAGES 168–177

1. Brown, Sanborn, *Benjamin Thompson, Count Rumford*, The MIT Press, Cambridge, MA, 1979.
2. Boerhaave, Herman, *The Elements of Chemistry: Being the Annual Lectures of, MD*, translated from the original Latin by Timothy Dallowe (1735).
3. Brown, *Benjamin Thompson, Count Rumford*.
4. Brown, G.I., *Scientist, Soldier, Statesman and Spy, Count Rumford*, Alan Sutton, Stroud, 2001, p. 120.
5. Cardwell, W to C, p. 98.
6. Vrest, Orton, *The Forgotten Art of Building a Good Fireplace*, Yankee Books, Emmaus, PA, 1974.
7. Brown, Sanborn, 'An experimental inquiry concerning the source of the heat which is excited by friction', in *Benjamin Thompson—Count Rumford: Count Rumford on the Nature of Heat*, Pergamon Press, Oxford, 1967, ch. 4, Essay IX, p. 55.
8. Ibid., p. 58.
9. Ibid., p. 60.
10. Ibid., p. 61.
11. Ibid., p. 64.
12. Ibid., p. 70.
13. Ibid., p. 70.
14. Brown, Sanborn, 'An inquiry concerning the weight ascribed to heat', in *Benjamin Thompson—Count Rumford*, ch. 6, p. 98.
15. Ibid., p. 90.
16. Rumford, Count, 'Of the propagation of heat in fluids', in Brown, Sanborn (ed.), *Collected Works of Count Rumford*, vol. 1, Belknap Press of Harvard University Press, Cambridge, MA, 1968, p. 190.
17. Davy, Humphry, 'Essay on heat and light', in Beddoes, Thomas, *Contributions to Physical and Medical Knowledge, Principally from the West of England*, printed by Biggs & Cottle, for T.N. Longman and O. Rees, Paternoster Row, London, 1799.
18. Andrade, E.N. da C., 'Two historical notes', *Nature*, cxxxv, 1935, p. 359.

19. Asimov, Isaac, quoted in Robinson, Andrew, *The Last Man Who Knew Everything*, Pi Press, Pearson Education, New York, 2006, p. 2.
20. Feynman, Richard, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1963, vol. I, ch. 37, end of 37–1.
21. DSB on Young, p. 567.
22. Ibid., p. 568, reference 64 therein.
23. Whittaker, Sir Edmund, *A History of the Theories of Aether and Electricity*, vol. 1: *The Classic Theories*, Harper & Bros, New York, 1960.
24. Robinson, *The Last Man Who Knew Everything*, p. 7.
25. Young, Thomas, *A Course of Lectures on Natural Philosophy and the Mechanical Arts*, 4 vols, Thoemmes, Bristol, 2003; a facsimile reprint of the original 1807 edition.
26. Davy, Humphry, First Bakerian Lecture, 'On some chemical agencies of electricity', 1806.
27. Cardwell, W to C, p. 112.
28. Davy, Humphry, *Elements of Agricultural Chemistry*, Longman, London, 1813.
29. Crosland, Maurice, *Gay-Lussac, Scientist and Bourgeois*, Cambridge University Press, Cambridge, 2004, p. 86.
30. Cardwell, W to C pp. 111–14.
31. As above, p. 105 and reference therein: Francis Trevithick, *The Life of Richard Trevithick*, Spon, London, 1874, p. 113.

CHAPTER 10, PAGES 178–199

1. Dalton, John, *A New System of Chemical Philosophy*, Part II, Manchester, 1810, 'Simple atmospheres and compound atmosphere'; in Greenaway, Frank, *John Dalton and the Atom*, Heinemann, London, 1966.
2. Holmyard, E.J., *Makers of Chemistry*, The Clarendon Press, Oxford, 1953, p. 222.
3. Dalton, John, Fourth 'Experimental Essay', entitled 'On the thermal expansion of gases', presented in 1800 and in *Memoirs of the Manchester Literary and Philosophical Society*, 1802.
4. Crosland, Maurice, *Gay-Lussac, Scientist and Bourgeois*, Cambridge University Press, Cambridge, 2004.
5. Cardwell, W to C, p. 131.
6. Crosland, *Gay-Lussac, Scientist and Bourgeois*.
7. Fox, Robert, 'The fire piston and its origins in Europe', *Technology and Culture*, 10(3), 1969, pp. 355–70.
8. Cardwell, W to C, p. 136.
9. Mendoza, E., Introduction to Sadi Carnot's *Reflections on the Motive Power of Fire*, Dover Publications, New York, 1960; and Cardwell, W to C, p. 137.

10. Fox, Robert, 'The background to the discovery of Dulong and Petit's Law', *British Journal for the History of Science*, 4(1), 1968, pp. 1–22.
11. Crosland, *Gay-Lussac, Scientist and Bourgeois*.
12. Fox, 'The background to the discovery of Dulong and Petit's Law'.

CHAPTER 11, PAGES 201–206

1. Herivel, J.W., 'Aspects of French theoretical physics in the nineteenth century', *British Journal for the History of Science*, 3(10), 1966, p. 112.
2. Biographical notes preceding Fourier, Joseph, *The Analytical Theory of Heat* (1822), Britannica Great Books, Vol. 45, 1952.
3. Herivel, 'Aspects of French theoretical physics', p. 113.
4. Fourier, Joseph, *The Analytical Theory of Heat* (1822), Britannica Great Books, vol. 45, 29th edn, 1987, preliminary discourse, p. 169.
5. Ibid., section I, first chapter, p. 177.
6. Ibid., preliminary discourse, p. 171.
7. Cardwell, W to C, ref. 64, p. 117.
8. Cardwell, W to C, p. 117.
9. Fourier, *The Analytical Theory of Heat*, preliminary discourse, p. 174.
10. Gillispie, Charles C., *Pierre-Simon Laplace, 1749–1827*, Princeton University Press, Princeton, NJ, 1997, p. 249.
11. DSB on Fourier, p. 97.
12. Biographical notes preceding Fourier, *The Analytical Theory of Heat*.
13. Cardwell, W to C, footnote on p. 146.
14. Herapath, John, 'Mr Herapath on the causes, laws and phaenomena of heat, gases etc.', *Annals of Philosophy*, 9, 1821, pp. 273–93; see p. 279.
15. Ibid., p. 282.
16. 'X', 'Remarks on Mr Herapath's Theory', *Annals of Philosophy*, ii, 1821, p. 390.
17. Cardwell, W to C.

CHAPTER 12, PAGES 209–226

1. Carnot, Sadi, *Réflexions sur la puissance motrice du feu* (*Reflections on the Motive Power of Fire*), translated by R.H. Thurston, edited with an introduction by E. Mendoza, Dover Publications, New York, 1960.
2. Ibid.
3. Ibid.
4. Ibid.
5. Ibid.
6. Ibid.
7. Ibid.
8. Ibid.

9. DSB on Lazare Carnot.
10. Ibid.
11. Carnot, *Réflexions*.
12. Gillispie, Charles C., quoted in Robert Fox's edition and translation of Sadi Carnot's *Réflexions sur la puissance motrice du feu*, Manchester University Press, Manchester, 1986.
13. Carnot, *Réflexions* (1960).
14. Ibid.
15. Ibid.
16. Ibid.
17. Ibid.
18. Ibid.
19. Ibid.
20. Ibid.
21. Ibid., footnote, p. 46.
22. Ibid., p. 46.
23. Ibid., 'Appendix: selections from the posthumous manuscripts of Carnot'.
24. Ibid.
25. Carnot, *Réflexions* (1986), p. 17.
26. Clapeyron, Émile, 'Memoir on the motive power of heat', in Carnot, *Réflexions* (1960); and in *Journal de l'École Polytechnique*, XIV, 1834, p. 153.
27. Thompson, Sylvanus, *The Life of William Thomson, Baron Kelvin of Largs*, Macmillan, London, 1910.
28. Carnot, *Réflexions* (1960).
29. Ibid., note on p. 40.

CHAPTER 13, PAGES 230–241

1. Green, George, *An Essay on the Mathematical Analysis of Electricity and Magnetism*, Nottingham, 1828; also in Green, George, *Mathematical Papers*, edited by N.M. Ferrers, Macmillan, London, 1871 (reprinted by Chelsea, New York, 1970).
2. St Andrew's University 'McTutor' biographies of mathematicians, at http://www-history.mcs.st-andrews.ac.uk/Mathematicians/Green.html
3. Cannell, D.M., *George Green, Mathematician and Physicist 1793–1841: the Background to His Life and Work*, Athlone Press, London, 1993.
4. Ibid., p. 170.
5. Ibid., p. 70.
6. Ibid., p. 146; also in Thompson, Sylvanus, *The Life of William Thomson, Baron Kelvin of Largs*, Macmillan, London, 1910, pp. 113–19.
7. Hankins, Thomas L., *Sir William Rowan Hamilton*, Johns Hopkins University Press, Baltimore, MD, 1980.

8. Ibid.
9. Bernoulli, Johann (1697), in *Opera Johannis Bernoullii*, vol. 1, pp. 206–11.
10. Hamilton, William R., 'On a general method of expressing the paths of light, and of the planets, by the coefficients of a characteristic function', *Dublin University Review and Quarterly Magazine*, I, 1833, pp. 795–826.
11. I am grateful to Murray Peake for this analogy and for useful discussions.
12. Hankins, *Sir William Rowan Hamilton*, p. 92, notes 15 and 16.
13. Ibid.
14. Kilmister, C.W., *Hamiltonian Dynamics*, Longman, London, 1964.
15. Hankins, *Sir William Rowan Hamilton*, p. 142.
16. Ibid., p. 164.
17. Ibid., p. 204.
18. Ibid., p. 64, note 7.

CHAPTER 14, PAGES 248–263

1. Lindsay, Robert Bruce, *Julius Robert Mayer, Prophet of Energy*, Selected Readings in Physics, General Editor D. ter Haar, Pergamon Press, Oxford, 1973, pp. 7–8.
2. Mayer, Julius R., 'Comments on the mechanical equivalent of heat' (1851), in Lindsay, *Julius Robert Mayer*, p. 203.
3. Mayer, Julius R., 'The motions of organisms and their relation to metabolism: an essay in natural science' (1845), in Lindsay, *Julius Robert Mayer*, p. 99.
4. Ibid.
5. Ibid., p. 74.
6. Ibid., p. 72.
7. Ibid., p. 86.
8. Cardwell, D.S.L., *James Joule, a Biography*, Manchester University Press, Manchester, 1989, p. 32.
9. Ibid., p. 45.
10. Ibid., p. 45.
11. Joule, James, 'On the calorific effects of magneto-electricity and on the mechanical value of heat', *Philosophical Magazine*, 23, 1843, pp. 263–76.
12. Ibid.
13. Ibid.
14. Joule, 'On the calorific effects', postscript; and Cardwell, *James Joule, a Biography*, p. 58.
15. Cardwell, *James Joule, a Biography*, p. 76.
16. Ibid., p. 61.
17. Ibid., p. 99, this and subsequent quotes.
18. Cardwell, *James Joule, a Biography*, footnote, p. 96.
19. DSB on Waterston, p. 185.

20. Ibid.
21. Ibid.
22. Cardwell, *James Joule, a Biography*, p. 99.
23. Ibid., p. 76.
24. Ibid., p. 43.
25. Joule, 'On the calorific effects'.
26. Cardwell, *James Joule, a Biography*, footnote, p. 96.
27. Ibid., p. 126.
28. Ibid., p. 208.

CHAPTER 15, PAGES 265–282

1. Pearce Williams, L., *Michael Faraday, a Biography*, Da Capo Press, New York, 1987, ch. 1, p. 28, ref. 47 (republication of original edition, Basic Books, New York, 1965).
2. Ibid., p. 29.
3. DSB on Faraday.
4. Quoted on the title page of Grove, W.R. (ed.), *The Correlation and Conservation of Forces*, containing essays by Grove, Helmholtz, Mayer, Faraday, Liebig and Carpenter, published by Appleton and Co., New York, 1867.
5. Pearce Williams, *Michael Faraday, a Biography*, p. 364.
6. Feynman, Richard, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1966, vol. II, end of section 17–1.
7. Pearce Williams, *Michael Faraday, a Biography*, p. 364.
8. Ibid., p. 257.
9. Ibid., p. 367.
10. Ibid., ch. 9, p. 403, note 9.
11. Ibid., p. 385.
12. Ibid., p. 386.
13. Ibid., p. 393.
14. Ibid., p. 70; and Agassi, Joseph, 'An unpublished paper of the young Faraday', *Isis*, 52, 1961, p. 88.
15. Pearce Williams, *Michael Faraday, a Biography*; and also Faraday's essay in Grove, *The Correlation and Conservation of Forces*.
16. Pearce Williams, *Michael Faraday, a Biography*, p. 466.
17. Ibid., p. 468.
18. Ibid., p. 468.
19. Ibid., p. 469.
20. Ibid., top of p. 470.
21. Faraday, Michael, 'Some thoughts on the conservation of force', in Grove, *The Correlation and Conservation of Forces*, p. 368.
22. M'Kendrick, J.G., *Hermann Ludwig Ferdinand von Helmholtz*, Longmans, Green & Co., New York, 1899, school report, p. 6.

23. Königsberger, Leo, *Hermann von Helmholtz*, translated by F.A. Welby, preface by Lord Kelvin, Dover Publications, New York, 1906, p. 17.
24. Ibid., p. 14.
25. Kahl, Russell (ed.), 'An autobiographical sketch', *Selected Writings of Hermann von Helmholtz*, Wesleyan University Press, Middletown, CT, 1971, ch. 17, p. 470.
26. Königsberger, *Hermann von Helmholtz*.
27. Kahl, 'An autobiographical sketch'.
28. Ibid.
29. Bevilacqua, F., in Cahan, David (ed.), *Hermann von Helmholtz and the Foundations of Nineteenth Century Science*, University of California Press, Berkeley, 1993, p. 300.
30. M'Kendrick, *Hermann Ludwig Ferdinand von Helmholtz*, p. 35.
31. Königsberger, *Hermann von Helmholtz*, ch. V.
32. Helmholtz, Hermann von, 'The conservation of force: a physical memoir' (1847), in Kahl, *Selected Writings of Hermann von Helmholtz*; all page numbers refer to this edition.
33. Ibid., p. 3.
34. Ibid., p. 7.
35. Ibid.
36. Ibid.
37. Ibid., p. 23.
38. Ibid., footnote 9, p. 20.
39. Ibid.
40. Ibid., p. 32.
41. Ibid.
42. Jungnickel, C. and McCormmach, R., *The Intellectual Mastery of Nature*, University of Chicago Press, Chicago, 1986, vol. 1, p. 159.
43. Helmholtz, 'The conservation of force: a physical memoir' (1847), in Kahl, *Selected Writings of Hermann von Helmholtz*, p. 44.
44. Ibid.
45. Darrigol, Olivier, *Electrodynamics from Ampère to Einstein*, Oxford University Press, Oxford, 2000.
46. Helmholtz, Hermann von, 'The conservation of force: a physical memoir', p. 43.
47. Ibid., p. 43.
48. Ibid., p. 43.
49. Ibid., p. 48.
50. Ibid., p. 48.
51. Ibid., p. 48.
52. Ibid., p. 49.
53. Kahl, 'An autobiographical sketch', p. 471.
54. Darrigol, *Electrodynamics from Ampère to Einstein*, p. 216.

55. Smith, Crosbie, *The Science of Energy*, The Athlone Press, London, 1998, p. 126.
56. Ibid., p. 126.
57. Maxwell, J.C., 'Hermann Ludwig Ferdinand Helmholtz', *Nature*, 15, 1877, pp. 389–91.
58. Königsberger, *Hermann von Helmholtz*.
59. Ibid.
60. Ibid.
61. Ibid.
62. Ibid.
63. Ibid.
64. Ibid.
65. Maxwell, 'Hermann Ludwig Ferdinand Helmholtz'.
66. Ibid.
67. Smith, *The Science of Energy*.
68. Heidelberger, M., 'Force, law and experiment', in Cahan, *Hermann von Helmholtz*, p. 180.
69. Helmholtz, 'The conservation of force: a physical memoir', p. 43.
70. Ibid., Appendix 4 added to memoir in 1881, p. 52.

CHAPTER 16, PAGES 285–303

1. Clapeyron, Émile, 'On the motive power of heat' (1834), in Eric Mendoza's edition of Sadi Carnot's *Reflections on the Motive Power of Fire*, Dover Publications, New York, 1960.
2. Thompson, Sylvanus, *The Life of William Thomson, Baron Kelvin of Largs*, Macmillan, London, 1910.
3. Thomson, William, 'An account of Carnot's theory of the motive power of heat, with numerical results deduced from Regnault's experiments on steam', *Transactions of the Royal Society of Edinburgh*, 16, 1849, pp. 541–74.
4. Clausius, Rudolf, 'On the moving force of heat and the laws of heat that may be deduced therefrom' (1850); translated by W.F. Magie, in Carnot, *Reflections* (1960).
5. Ibid.
6. Ibid.
7. Ibid.
8. Thomson, William, 'On the dynamical theory of heat, with numerical results from Mr Joule's equivalent of a thermal unit and M. Regnault's observations on steam', *Transactions of the Royal Society of Edinburgh*, March 1851, this and following quotation.
9. Carnot, Sadi, Appendix to *Réflexions sur la puissance motrice du feu*, translated by R.H. Thurston, edited with an introduction by E. Mendoza, Dover Publications, New York, 1960.

10. Thomson, William, 'Thermo-electric currents, preliminary 97–101, Fundamental principles of general thermo-dynamics recapitulated', *Transactions of the Royal Society of Edinburgh*, vol. xxi, part I, p. 232, read May 1854; in *Mathematical and Physical Papers*, vol. 1, Cambridge University Press, Cambridge, 1882.
11. Smith, Crosbie and Wise, Norton M., *Energy and Empire: a Biographical Study of Lord Kelvin*, Cambridge University Press, Cambridge, 1989.
12. Thomson, William, 'On a Universal Tendency in Nature to the Dissipation of Mechanical Energy', *Philosophical Magazine*, 1852.
13. Cardwell, W to C, p. 272.
14. Ibid., p. 269.
15. Smith and Wise, *Energy and Empire*.
16. Thomson, William and Tait, Peter G., *Treatise on Natural Philosophy*, Macmillan for the University of Oxford, 1867.
17. Smith and Wise, *Energy and Empire*.
18. Brush, Stephen G., *The Temperature of History: Phases of Science and Culture in the Nineteenth Century*, Lenox Hill, 1979.

CHAPTER 17, PAGES 304–323

1. Grattan-Guiness, Ivor, 'Work is for the workers: advances in engineering mechanics and instruction in France, 1800–1830', *Annals of Science*, 41, 1984, pp. 1–33.
2. Knudsen, Ole, 'Electromagnetic energy and the early history of the energy principle', in Kox, A.J. and Siegel, D.N. (eds), *No Truth Except in the Details: Essays in Honor of Martin J. Klein*, Kluwer Academic, Dordrecht, 1995.
3. Gillispie, Charles C., *Essays and Reviews in History and History of Science*, American Philosophical Society, Philadelphia, PA, 2007, ch. 19.
4. Quetelet, Adolphe, *A Treatise on Man and the Development of His Faculties*, William and Robert Chambers, Edinburgh, 1842.
5. Maxwell, J.C., *Theory of Heat*, 1871 and 1888, introduction by Peter Pesic, Dover Publications, New York, 2001, p. 329. (He wrote 'the equality which we assert to exist between the molecules of hydrogen applies to each individual molecule, and not merely to the average of groups of millions of molecules'.)
6. Müller, Ingo, *A History of Thermodynamics: the Doctrine of Energy and Entropy*, Springer-Verlag, Berlin, 2007, p. 92.
7. Klein, Martin J., 'The development of Boltzmann's statistical ideas', in Cohen, E.G.D. and Thirring, W. (eds), *The Boltzmann Equation: Theory and Applications*, Springer-Verlag, Vienna, 1973.
8. Ibid.
9. Ibid.
10. Ibid.
11. Ibid.

12. Ehrenfest, Paul and Ehrenfest, Tatyana (1911), English translation in Moravcsik, M.J., *The Conceptual Foundations of Statistical Mechanics*, Cornell University Press, Ithaca, NY, 1959.
13. Gibbs, Josiah Willard, 'On the equilibrium of heterogeneous substances', in Bumstead, H.A. and Van Name, R.G. (eds), *The Scientific Papers of J. Willard Gibbs*, Longmans, Green & Co., London, 1906; reprinted by Dover Publications, New York, 1961.
14. Maxwell, James Clerk, in Gillispie, Charles C., *The Edge of Objectivity*, Princeton University Press, Princeton, NJ, 1973.
15. Feynman, R.P., *Lectures on Physics*, with R. Leighton and M. Sands, Addison-Wesley, Reading, MA, 1966, vol. II.
16. Feynman, R.P., 'The space-time formulation of nonrelativistic quantum mechanics', *Reviews of Modern Physics*, 20, 1948, pp. 367–87; and Feynman, R.P. and Hibbs, A.R., *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, 1965.
17. Einstein, Albert, in Klein, Martin J., 'Thermodynamics in Einstein's thought: thermodynamics played a special role in Einstein's early search for a unified foundation of physics', *Science*, 157, 1967, pp. 509–16.
18. Ibid.
19. Rindler, Wolfgang, *Special Relativity*, 2nd edn, Oliver & Boyd, Edinburgh/ Interscience Publishing, New York, 1966, p. 88.
20. Einstein, Albert, in Pais, A., *Subtle is the Lord: the Science and the Life of Albert Einstein*, Oxford University Press, Oxford, 2005.
21. Feynman, R.P., The reason for antiparticles, in *Elementary Particles and the Laws of Nature: the 1986 Dirac Memorial Lectures*, with S. Weinberg, Cambridge University Press, Cambridge, 1987.
22. Cardwell, D.S.L., *James Joule, a Biography*, Manchester University Press, Manchester, 1989, p. 253.
23. Ibid., p. 251.
24. Ibid.
25. Thomson, William and Tait, Peter G., *Treatise on Natural Philosophy*, Macmillan for the University of Oxford, 1867.
26. Smith, Crosbie and Wise, Norton M., *Energy and Empire: a Biographical Study of Lord Kelvin*, Cambridge University Press, Cambridge, 1989.
27. Maxwell, J.C., *Theory of Heat* (1891), Dover Publications, New York, 2001, Introduction and notes by P. Pesic.
28. Smith, Crosbie, *The Science of Energy: a Cultural History of Energy Physics in Victorian Britain*, The Athlone Press, London, 1998, p. 260.
29. Ibid., p. 258.
30. Cercignani, Carlo, *Ludwig Boltzmann: the Man Who Trusted Atoms*, Oxford University Press, Oxford, 1998, p. 200.
31. Lindley, David, *Boltzmann's Atom: the Great Debate that Launched a Revolution in Physics*, The Free Press, New York, 2001, p. 81.

32. Dyson, Freeman, ‘Why is Maxwell’s Theory so Hard to Understand?’; available at http://www.clerkmaxwellfoundation.org/DysonFreemanArticle.pdf.
33. Ibid.
34. Brush, Stephen G., *The Temperature of History: Phases of Science and Culture in the Nineteenth Century*, Lenox Hill, 1979, p. 61.
35. Ibid., p. 40.
36. Smith and Wise, *Energy and Empire*, p. 317.
37. Ibid., p. 500.
38. Müller, *A History of Thermodynamics*.
39. Brush, *The Temperature of History*, pp. 63 and 64.
40. Ibid.
41. Ibid.
42. Ibid., p. 43.
43. Smith and Wise, *Energy and Empire*, p. 566.
44. Smith, *The Science of Energy*, p. 251.
45. Cercignani, *Ludwig Boltzmann*, p. 84.
46. Lindley, *Boltzmann's Atom*, Introduction.
47. Cercignani, *Ludwig Boltzmann*, p. 202.
48. Kac, Mark, Probability and related topics in physical sciences, in Summer Seminar in Applied Mathematics, Boulder, Colorado, 1957, published as *Lectures in Applied Mathematics*, vol. 1.1, American Mathematical Society, 1957.
49. Lindley, *Boltzmann's Atom*, p. 102.
50. Ibid., p. 225.
51. Gibbs, ‘On the equilibrium of heterogeneous substances’.
52. Angrist, S. and Hepler, L., *Order and Chaos*, Penguin, London, 1973.
53. DSB on Clausius.
54. Cercignani, *Ludwig Boltzmann*, p. 176.
55. Coopersmith, Bertie, personal communication.
56. Cardwell, *James Joule, a Biography*, p. 216.
57. Ibid., p. 259.
58. Ayrton, E., *Nature*, 72, 1905, p. 612.

CHAPTER 18, PAGES 324–349

1. Pippard, Brian, *The Elements of Classical Thermodynamics*, Cambridge University Press, Cambridge, 1957.
2. Zemansky, Mark, *Heat and Thermodynamics*, 5th edn, McGraw-Hill Kogakusha, Tokyo, 1968.
3. Zemansky, *Heat and Thermodynamics*, p. 1.
4. Ibid., p. 78 (see also Pippard, *The Elements of Classical Thermodynamics*, p. 16).

5. Zemansky, *Heat and Thermodynamics*, p. 71.
6. Pippard, *The Elements of Classical Thermodynamics*, p. 15.
7. Zemansky, *Heat and Thermodynamics*, p. 76.
8. Pippard, *The Elements of Classical Thermodynamics*, p. 9.
9. Ibid., p. 12.
10. Zemansky, *Heat and Thermodynamics*, p. 148.
11. Ibid., p. 433.
12. Coopersmith, Bertie, personal communication.
13. Feynman, Richard, *Lectures on Physics*, with R. Leighton and M. Sands, Addison-Wesley, Reading, MA, 1966, vol. II.
14. Planck, Max, *Treatise on Thermodynamics*, 1905, translated by Alexander Ogg, reprinted by Dover Publications, New York, 1990, pp. 106–7.
15. Poincaré, Henri, in Doughty, Noel, *Lagrangian Interaction*, Addison-Wesley, Reading, MA, 1990, p. 225.
16. Pauli, Wolfgang's letter in Cropper, W.H., *Great Physicists: the Life and Times of Leading Physicists from Galileo to Hawking*, Oxford University Press, Oxford, 2001, p. 335.
17. Ibid.
18. This example is modified from the one in Duffin, W.J., *Electricity and Magnetism*, 2nd edn, McGraw-Hill, New York, 1973, p. 375.
19. Maxwell, James Clerk, *Theory of Heat*, 1871; Dover Publications, New York (2001) new introduction and notes by Peter Pesic; see p. 301.
20. Ibid., p. 302.
21. Jeans, James, *An Introduction to the Kinetic Theory of Gases*, Cambridge University Press, Cambridge, 1940; reissued 2008.
22. Collie, C.H., *Kinetic Theory and Entropy*, Longman, London, 1982, section 5.2, p. 159; also Zemansky, *Heat and Thermodynamics*, section 4-6, p. 82.
23. Pippard, *The Elements of Classical Thermodynamics*, p. 78.
24. Ibid., p. 78.
25. Feynman, *Lectures on Physics*, vol. I, section 39-4.
26. Pippard, *The Elements of Classical Thermodynamics*; and Zemansky, *Heat and Thermodynamics*.
27. Klein, Martin J., 'The development of Boltzmann's statistical ideas', in Cohen, E.G.D. and Thirring, W. (eds), *The Boltzmann Equation: Theory and Applications*, Springer-Verlag, Vienna, 1973.
28. Gibbs, Josiah Willard, 'On the equilibrium of heterogeneous substances', in Bumstead, H.A. and Van Name, R.G. (eds), *The Scientific Papers of J. Willard Gibbs*", Longmans, Green & Co., London, 1906; reprinted by Dover Publications, New York, 1961.
29. Pippard, *The Elements of Classical Thermodynamics*, p. 30 (Pippard writes 'adiathermal' instead of 'adiabatic').
30. Müller, Ingo, *A History of Thermodynamics: the Doctrine of Energy and Entropy*, Springer-Verlag, Vienna, 2007, p. 323.

CHAPTER 19, PAGES 351–359

1. Brush, Stephen G., *The Kind of Motion We Call Heat*, North-Holland, 1976, Book 1, p. 44.
2. Poussin, Nicolas, 'A Dance to the Music of Time', oil on canvas, *c*. 1640, Wallace Collection, London.
3. Hankins, Thomas L., *Sir William Rowan Hamilton*, Johns Hopkins University Press, Baltimore, MD, 1980, p. 177.
4. Coopersmith, Jennifer, 'Yes, we really *know* something', in preparation.
5. Howard Doherty of Doherty's Garage, St Andrews Street, Bendigo, personal communication.

Index